

Malaria after the genome

Following the publication of the malaria parasite's genome sequence and the beginnings of relevant proteomics, research tools are now available that could make a big difference in the long-term war against malaria.

Tuberculosis. AIDS. Malaria. The three great modern plagues of the world's poor. The battle against them now stands at a crucial juncture. A committed effort by governments will determine whether, or how soon, victory is to be had. The United States in particular has a key role to play in shaping the current transformation in the way the world deals with international health, in the largest shift since the World Health Organization (WHO) was created as part of the United Nations after the Second World War.

The WHO retains its key role as the broker of intergovernmental relations between the health ministries of 192 countries. But over the past decade, a global reawakening of consciousness that these scourges can and must be defeated has spawned a plethora of new international players. These include collaborations between the public and private sectors, initiatives by charities, and the Global Fund to Fight AIDS, Tuberculosis and Malaria, which is yet to prove itself but is perhaps the first realistic effort to control these diseases using existing tools.

The United States should bring leadership to this movement. In particular, it should ensure that existing science is used effectively, by supporting control efforts through more generous support for bodies such as the Global Fund. It is widely agreed that a proper control programme for malaria would cost at least US\$2 billion annually — the equivalent of just two days' worth of US and European Union farm subsidies. It is less a question of money than of political will.

But existing science cannot on its own control malaria or other neglected diseases of the poor. We need science to yield new tools. In particular, there is a golden opportunity to engender solid progress in the fight against malaria, which every day kills between 2,500 and 8,000 children under five years of age. That we do not know how many children die says much about the lack of demographic and epidemiological research. One-sixth of the Earth's population suffers from the disease, crippling economic growth in those countries across South America, Africa and Southeast Asia that need it most. The challenges are made all the more urgent by the development of multidrug resistance by the parasite (see *Nature* **415**, 670–672; 2002).

On the map

Science has now provided such an opportunity. *Nature* recently published the genome sequence of the malaria parasite *Plasmodium falciparum*, and with it, fundamental information that is critical to understanding the parasite. Whereas malaria researchers previously had to slog for months or years to find and study this or that gene or protein, a consortium of US and other researchers has now made maps of them all available to the world. And that for the modest sum of \$15 million over six years.

This is an important step along the way, and provides a new platform on which to develop essential insights into the many bio-molecular and cellular pathways by which the parasite survives and interacts with us, its indispensable host. New techniques are beginning to be applied, such as high-throughput analyses of the pattern of gene expression and of the interactions of proteins at key phases of the parasite's life cycle. The availability of the genome combined with these techniques will undoubtedly spur progress significantly — if there are funds and the international structures to permit it.

A step towards the latter is the Multilateral Initiative on Malaria, created in 1997 to coordinate worldwide research, in large part thanks to the vision of Harold Varmus, then director of the US National Institutes of Health (NIH). The funds needed would not be enormous in absolute terms for the science that would flow from a post-genomics initiative. But they would be significant relative to the total world funding for malaria research, currently around \$100 million annually.

International effort

Inevitably, detractors will argue that post-genomics research is a luxury, and that money would be better spent on developing better drugs, or a vaccine, or on enhancing control of the disease through prevention. This misses the point. Funds for malaria research and drug and vaccine development are woefully inadequate (see *Nature* **419**, 426–428; 2002). But history shows that every one of these alternative routes has its own chronic difficulties. Addressing new opportunities in the basic science of the disease should, in the long term, deliver far-reaching solutions. This should be funded by new money, increasing the pool of malaria research funds. It would therefore be widely recognized as a wonderfully enlightened action if the NIH were to take the lead in creating a substantial international malaria post-genomics programme.

An international effort would be money well spent. The availability of the genome has already stimulated lightning progress in deciphering the parasite's proteome by attracting to malaria research some of the brightest minds from the yeast community. There are many excellent researchers who would make rapid progress in malarial post-genomics if substantial new money were made available.

The malaria genome is already providing a service to the wider malaria research community. African scientists are using molecular markers of drug resistance, giving rise to country-wide 'early warning' maps; resistance was previously measured when it was too late, by the number of people who failed to respond to treatment. A post-genomics effort would yield valuable data and reagents for scientists in almost every area of malaria research.

Any new effort should pay particular attention to dissemination, through workshops, training and online databases, in particular to benefit scientists in Africa, where nine out of ten deaths from malaria occur. It also needs to guard against becoming a machine solely for generating scientific papers, disconnected from the wider war on the disease. Malaria post-genomics should not be an island, but should be as inclusive as possible. It should also be firmly anchored to existing nascent initiatives to ensure as much cross-fertilization as possible among areas of research.

For the United States to provide significant new help in this battle would not simply represent an act of great goodwill. It would also be in the nation's long-term strategic interests. Reducing developing countries' battle with the illness and mortality of malaria and the social burdens that they bring would allow them to focus more on economic and social development, providing new opportunities for US businesses and other organizations. It's over 30 years since humans set foot on the Moon; the time has come for a decade-long initiative towards freeing mankind from malaria. ■





War on science
Research could be hit
by weak US economy
and military plans
p100



Gritted teeth
China seeks the
source of Beijing's
dust storms
p101



Square dance
Project director sets
out plans for radio
telescope array
p103



On the double
Gold anniversary
sees DNA strike a
pose on £2 coin
p103

Los Alamos directors resign after row over sacked whistleblowers

Geoff Brumfiel, Washington

Accusations of mismanagement and theft at the United States' most venerable nuclear-weapons laboratory have prompted the resignations of its top two managers.

On 2 January, John Browne and Joseph Salgado stepped down as director and deputy director of the Los Alamos National Laboratory in New Mexico. Their resignations are believed to have been requested by the University of California (UC), which has overseen the laboratory for the US government since 1943.

Congressional and Department of Energy officials are now demanding a review of the university's contract to manage the lab, whose once-formidable reputation has been tarnished by a succession of recent crises.

The resignations stemmed from allegations made in November, when two independent investigators hired by the lab released documents that they said showed endemic theft and credit-card fraud at Los Alamos. Laboratory officials said the alleged problems amounted to only a few million dollars, against the lab's annual budget of \$1.3 billion for goods and services. Shortly afterwards, the lab fired the investigators, without giving any public explanation.

This angered the congressional and federal officials who oversee the lab. "The treatment of whistleblowers is a big issue for us," said a staff member on a congressional committee investigating the situation at the lab. This sentiment was strongly echoed by energy secretary Spencer Abraham in a letter sent to Richard Atkinson, the UC president, on 24 December, in which he said that the two investigators' dismissals were "of most immediate concern". Abraham had already dispatched the energy department's inspector general to look into the matter.

The incident is the latest in a string of embarrassing mishaps for the lab, which was the birthplace of the atomic bomb. In 1999, Los Alamos scientist Wen Ho Lee was accused of passing nuclear secrets to a foreign power (see *Nature* 398, 96; 1999); he was later acquitted. In 2000, computer drives containing secret bomb data disappeared from a secure vault and later reappeared



Bombshell: John Browne's resignation as director of Los Alamos follows of a series of setbacks at the lab.

behind a photocopier (see *Nature* 405, 725; 2000). Earlier that year, a nearby forest fire had threatened sensitive areas of the lab.

"A whole string of things have been happening over the past few years," says Robert Civiak, a consultant who until 1999 worked as White House budget examiner for the

energy department's national security programmes. "This could just be the straw that breaks the camel's back."

The series of incidents caused the energy department to put stipulations in its January 2001 contract with UC aimed at bolstering supervision and security at the lab. Chief

Insurers left reeling by disaster year

Quirin Schiermeier, Munich

Natural disasters caused record economic damage worldwide of US\$55 billion last year — \$20 billion more than in 2001 — according to an analysis by reinsurance company Munich Re.

The annual survey by Munich Re, which provides policies to back insurance companies, helps climate researchers to match their own predictions against what is actually going on. No one has conclusively tied any increase in the rate of extreme weather events to global warming — but Munich Re analysts predict that global warming will increasingly cause such events.

August's floods in central Europe were the most expensive disaster of the year (see *Nature* 418, 905; 2002), with damage costing some \$18.5 billion, says the company.

But in terms of deaths and injuries, 2002

wasn't such a bad year. About 11,000 people were killed in natural disasters, compared with 25,000 in 2001, when earthquakes hit El Salvador and India.

The company's geoscience research group has been monitoring natural disasters for 30 years, using information from news agencies, the Red Cross and its own branches and clients in 150 countries.

The most notable trend during 2002 was the increase in extreme weather incidents, says Thomas Loster, head of climate risk research at Munich Re. "These observations are consistent with predictions made by the Intergovernmental Panel on Climate Change about more frequent weather extremes," says Loster. "Although single observations are not statistically significant, each is an important element of the climate mosaic." ■

♦ www.munichre.com/default_e.asp

among these was the appointment of a head of laboratory management for all the labs administered by the university. Retired Ford Motor Company vice-president John McTague was appointed, but resigned in November after trying to appoint a Los Alamos scientist to head the other main scientific laboratory in the US nuclear-weapons programme, the Lawrence Livermore National Laboratory in California. That appointment was overturned by the White House after strenuous objections from Livermore researchers (see *Nature* **417**, 3; 2002).

In his 24 December letter, Abraham said that the energy department would conduct a review of UC's contract to manage the lab, which would be completed by May. "There's more reason now than ever for the energy department to consider a body other than the University of California," says Ray Kidder, a retired nuclear-weapons scientist from Lawrence Livermore. But Kidder thinks that a change in the contract would damage the morale of scientists at Los Alamos.

Possible contenders for the contract include the University of Texas, which expressed interest in it in 2001, and the Battelle Corporation of Columbus, Ohio, which runs other energy department labs, including the Oak Ridge National Laboratory in Tennessee and (jointly with Stony Brook University) the Brookhaven National Laboratory in New York.

Vice-Admiral George Nanos, who previously ran the Threat Reduction Directorate at Los Alamos, will serve as the lab's interim director.

US societies unite in plea for boost to research budgets

Geoff Brumfiel, Washington

Science lobbyists have launched a last-ditch attempt to win funding increases for research in President Bush's 2004 budget proposal. But the weak economic outlook and possibility of war mean it could be a disappointing year for US science agencies.

On 27 December, 32 societies signed a letter to Bush and his budget director, Mitch Daniels, urging them to increase research funding in their budget, which will be released on 4 February. "We strongly urge you to increase support for science programmes," says the letter, which cited a string of statements from administration officials and advisers pledging improved science funding.

Earlier in 2002, it points out, the President's Council of Advisors on Science and Technology called for more support for research in the physical sciences (see *Nature* **419**, 3; 2002). And on 19 December, President Bush signed into law an act authorizing a doubling of the budget for the National Science Foundation (NSF) over five years.

But all the indications are that next month's budget won't deliver on these pledges. Because Congress has yet to finalize the budget for the 2003 fiscal year — which began last October — the Bush administration will use its own 2003 budget proposal as a guideline for next year's funding levels. And in some cases, those 2004 numbers are less than what Congress might appropriate for



Mitch Daniels is being urged to allot more cash to science.

the agencies in 2003. For example, one congressional source says that Bush will propose an NSF budget of about \$5.4 billion in 2004. This is 9% more than he proposed in 2003 — but actually no more than Congress is planning to give the agency this year.

Other agencies may face similar woes. Several reports suggest that the National Institutes of Health could win an increase of as little as 0.3% to its \$27-billion budget. And a Pentagon source says that laboratories in the defence department are fighting to preserve their fundamental research budgets in the face of a possible war with Iraq.

Meanwhile, the lack of a finalized budget meant that research agencies rang in the New Year still stuck at 2002 budget levels. But it is likely that their 2003 budgets will be agreed in the next few weeks (see *Nature* **419**, 657; 2002).

Given the uncertainty of the year ahead, "people have yet to believe that the administration is committed to increasing science funding", says Samuel Rankin, head of government relations at the American Mathematical Society. "It's sort of like 'the cheque's in the mail'."

British chemists warned of impending stagnation

David Adam, London

British chemistry has lost its cutting edge, says a major review of its performance, and now tends to follow where others lead.

The review — conducted by an international panel for the Engineering and Physical Sciences Research Council (EPSRC) — says that chemistry in Britain is "relatively conservative" compared with work in other countries, or even with its own past.

The panel, which was chaired by Harvard chemist George Whitesides, did reach some positive conclusions: the quality of British scholarship is comparable to the world's best, it finds, as are facilities at the top universities. But the chemistry community has failed to embrace multidisciplinary areas of research, such as materials science and chemistry at the interface with biology, it concludes.

"We believe the United Kingdom will benefit if the academic chemistry community becomes more innovative," says

the panel's report, *Chemistry at the Centre*.

The findings will make uncomfortable reading for British chemists, who are already facing a sharp decline in the number of undergraduate students who are entering

the discipline (see *Nature* **416**, 777; 2002). "We hope this serves as a wake-up call," says one panel member, who did not want to be identified, "because British chemistry is in danger of becoming marginalized."

The panel says that one reason many British academic chemists are less concerned with innovation is that they tend to have close ties with the mature chemical industry. It adds that the current funding system cannot support long-term, focused programmes, and that the discipline fails to attract sufficient recruits from overseas.

David Clark, director of research and innovation at the EPSRC, which funds most chemistry in British universities, agrees that "people are playing it safe" in the discipline. The agency will discuss the report at an open meeting at the Geological Society in London on 27 January, and will then draw up an action plan to update its strategy for supporting chemistry, he says.



China plans clean sweep on dust storms

David Cyranoski

Chinese researchers at home and abroad are rallying allies around the world to confront a gritty subject that has growing implications for public health and the environment: dust.

The dust storms, which arise in China's northwestern plains and normally hit Beijing each spring, have become increasingly intense over the past few years. And the country's leaders are concerned not just for the health of the capital city's 13 million residents, but also about the implications for the government's pledge to stage a 'green' Olympic Games there in 2008.

So China is pouring substantial funds into research on the origins and impact of the storms — equivalent to tens of millions of US dollars over the next few years, according to one leading researcher's estimate — with agencies and institutions scrambling for a slice of the action. "Dust-storm research is a golden baby in China — everyone wants to hold it," says Zhanqing Li, who uses satellite imaging to study the storms at the University of Maryland in College Park.

Li even predicts that dust-storm research could become the first field of modern science in which China assumes clear global leadership. Last November, the Chinese Academy of Sciences gathered scientists from nine countries — including France, Japan, Korea, Russia and the United States — in Beijing to discuss the integration of different countries' research on dust storms. The 62 delegates signed up to the Beijing Declaration on an International Dust Storm Program, under which the Chinese Academy of Sciences will "formulate the framework for

the development and implementation of an integrated scientific research program".

A Chinese-US workshop on dust storms and their effects on human health also took place in November at North Carolina State University in Raleigh. It produced an agreement on how to study dust storms, from their origins right through to their environmental and health impacts, says the meeting's organizer, Lian Xie, a climate modeller at North Carolina State. The workshop was initiated by the North Carolina chapter of the US-based Chinese Association for Science and Technology and the China Meteorological Administration, and drew sponsorship from five US government agencies. China regards collaboration with the United States as vital — particularly because of the latter's immense satellite-imaging capability.

Researchers from the two countries discussed common strategies for approaching all aspects of the dust problem. The North Carolina meeting highlighted the keen involvement of US-based, Chinese-born researchers in the subject. "Many of us lived through dust storms in China," Xie explains.

Dust storms, as they pass between continents, have been found to contain a cocktail of natural and industrial particulates and aerosols. "You get an incredible mix of all kinds of aerosols that you don't see anywhere else," says Joseph Prospero, an aerosol chemist at the University of Miami.

In China, meanwhile, many believe that desertification — which is partly the result of overgrazing and poor land-use policy — is a principal cause of the storms' increasing intensity. China has already relocated farmers in some areas in an attempt to address the



Choked: the health effects of Beijing's annual dust storms remain a clouded issue.

REUTERS

problem. But Prospero thinks it is more likely that long-term climate trends are to blame.

For some researchers, a question mark remains over governments' willingness to confront the effects of dust storms on human health. According to literature distributed at the North Carolina meeting, "noxious aerosols kill 1 in 14 people, and pneumonia, often from breathing dirty air, is the leading cause of death among children in China". Eugene Shinn, a geologist with the US Geological Survey in St Petersburg, Florida, says: "The problem keeps getting swept under the rug."

Indian prime minister pledges to revamp science

K. S. Jayaraman, Bangalore

India is to mount a determined effort to attract its scientists home from abroad and to strip its scientific agencies of excessive bureaucracy, under a science policy document released last week.

The *Science and Technology Policy 2003* — which has taken years to produce — was unveiled on 3 January by Prime Minister Atal Bihari Vajpayee at the Indian Science Congress in Bangalore. He appealed to India's "scientific diaspora" to return to the country to help him realize his "vision of making India a developed nation".

"We have to ensure that our scientific institutions do not become afflicted with the culture of our government agencies," Vajpayee said. "Seniority should not displace merit; talent should not be suppressed."

Vajpayee also pledged that the government would make the required budgetary commitment to increase Indian spending on research and development — by government and industry — to at least 2% of the gross domestic product by 2007. Valangiman Ramamurthi, the science and technology secretary, claims that this will not be difficult, as spending has already risen from 0.8% in 2000 to 1.08% in 2002.

Ramamurthi adds that "the mechanisms will be in place very soon" to attract home scientists who have left India. He says that the new policy will be rapidly implemented and will give universities and research institutions greater autonomy.

Government officials say that under the policy, science-based ministries will be run by scientists and engineers, and other

ministries will appoint scientific advisory committees. They also say that selected universities and scientific institutions will get money to strengthen their infrastructure. But details of the funding will be left to a task force being set up to find ways of encouraging private and public investment in research.

Scientists at the conference were encouraged by the promises of greater autonomy and scientific input in decision-making, but were sceptical about pledges on government reform. "If the prime minister thinks de-bureaucratization is needed only in science departments, he is wrong," says J. Gowrishankar, a molecular biologist at the Centre for DNA Fingerprinting and Diagnostics in Hyderabad. "It is needed in the finance ministry, which audits the scientific institutes."

NSF urged to take multidisciplinary tack on environment

Tony Reichhardt, Washington

The National Science Foundation (NSF) should build up environmental research in the United States by backing more multidisciplinary work and putting more resources into infrastructure and the training of young scientists, says an advisory committee.

The panel's 10-year-plan for environmental science at the NSF echoes earlier reports from both the National Academy of Sciences and the National Science Board (NSB) that governs the foundation. But the committee, which was chaired by Stephanie Pfirman, an environmental scientist at Barnard College in New York, failed to endorse a call by the NSB for a billion-dollar increase in annual environmental research spending by the NSF.

In 1999, the science board recommended tripling the NSF's \$600-million environmental research budget within five years. But since then, the programme has expanded only gradually, to \$829 million last year.

An academy report in 2000 identified eight "grand challenges" for environmental science, from understanding biodiversity to improving hydrologic forecasting. The advisory panel review, explains Margaret Leinen, an oceanographer who heads the NSF's geosciences directorate and coordinates its environmental research, was commissioned to recommend what part of that portfolio the NSF should take on between now and 2012.

The review says that the NSF should focus on three areas: coupled human and natural systems, coupled biological and physical systems, and people and technology. Leinen says that the breadth of these categories reflects the need for a multidisciplinary approach — and that the NSF's support for environmental research will cut sharply across the agency's traditional, discipline-based organization. Study of the environmental impact of urban development, for example, may well involve both the geosciences and the social and behavioural sciences directorates.

Such collaboration is feasible in the short term in some areas, such as building computer systems to mine environmental data, Leinen says. But other areas, she adds, including the interaction of environmental and social sciences, may take longer to develop. ■

Biologists seek blueprint for international stem-cell effort

Erika Check, Washington

Scientists and policy-makers from around the world were this week set to lay the foundations of an international collaboration in human stem-cell research.

Representatives from the United States, Canada, Australia, Singapore, Finland, Sweden and Israel were due to meet in London as *Nature* went to press at the invitation of George Radda, chief executive of Britain's Medical Research Council, to discuss ways to hasten the development of any therapies that might result from stem-cell research.

Supporters say that what they term a 'human stem-cell project' would help to ensure that scientists from countries where such research is restricted could still contribute to this emerging field of biology.

Rules for research using human embryonic stem cells vary around the world. In particular, there are divergent approaches to the derivation of new stem cell lines — a procedure that is politically controversial as it involves the use of human embryos for research purposes.

In the United States, researchers are not allowed to use federal funds to derive and study new stem-cell lines, and in Germany, the derivation of new lines is forbidden by law. Meanwhile, scientists in the United Kingdom, Israel and Singapore can create and work on new stem-cell lines.

Roger Pedersen, a developmental biologist who in 2001 left the University of California, San Francisco, for the University of Cambridge, UK, says that he expects the international consortium to help researchers to contribute to stem-cell research while abiding with this patchwork of laws.

"It would be realistic to expect that novel



George Radda: convening experts in London.

embryonic stem-cell lines would be derived in the United Kingdom, Sweden and Singapore, and that US efforts would tend to focus on understanding how to control the *in vitro* differentiation of existing human embryonic stem-cell lines," Pedersen says. He adds that the international consortium is expected to lay out its goals in a series of meetings during the first half of this year.

An international stem-cell consortium could prove politically sensitive, and Radda's staff said that he couldn't comment on the planned meeting. But policy-makers initially seemed to welcome the idea.

For example, James Battey, head of the stem-cell task force at the US National Institutes of Health (NIH), says that a central coordinating body could do much to advance stem-cell research around the world. He was planning to attend this week's meeting on behalf of the NIH. "By exploiting the strengths of what's possible in each country, maybe we'll be better able to move the science forward," Battey says. ■

Comet mission hangs in the balance

David Adam, London and Declan Butler, Paris

The European Space Agency's Rosetta mission faces an uncertain future this week after an agency inquiry failed to give it clearance for take off.

Rosetta, a US\$700-million plan to land a probe on the comet Wirtanen, was originally scheduled to launch on 12 January. But it was postponed after the failure of the Ariane 5 ECA 'heavy lifter' on its maiden flight last month (see *Nature* 420, 723; 2001).

An inquiry into whether the accident has any bearing on the classic Ariane 5 design, which will be used to take Rosetta on the first leg of its journey, reported on 7 January but

failed to give the mission a green light. Rosetta has a very narrow launch window and must blast off by 31 January if it is to meet the comet.

At a press conference in Paris, Jean-Yves Le Gall, chief executive of Arianespace, which operates the Ariane rockets, said that the accident was caused by a cooling fault on the new engine, which suggests that the older Ariane 5 model is not at risk. But there are still doubts over whether the fault will affect Ariane 5 under the unusual conditions of the Rosetta launch, which requires the rocket to enter an Earth orbit and then re-fire. A separate investigation is now checking this, and will report on 14 January. ■

Applied Biosystems blames staff layoff on lack of demand

San Francisco The lab-technology company Applied Biosystems says that falling demand for its gene-sequencing technologies is to blame for its decision to cut 500 jobs, mainly in the United States and Europe.

The cuts, which will be implemented this month, amount to 9% of the company's workforce. The redundancies follow the axeing of 132 jobs last June from the firm's affiliate company Celera Genomics, based in Rockville, Maryland, which has shifted its emphasis from selling genetic data to discovering new drug targets.

Applied Biosystems, based in Foster City, California, cites several reasons for the slump, including spending cuts in the pharmaceutical industry and the ongoing lack of a 2003 budget for the US National Institutes of Health.

India joins dissenters in row over transgenic food aid

New Delhi Arguments over the use of genetically modified crops in food aid have spread from Africa to India.

The issue hit the headlines last year, when Zambia refused to accept aid containing transgenic food. Now it has emerged that last May, India rejected a US shipment of maize and soya bound for the southern state of Orissa. The decision became public last week when Senator Kit Bond (Republican, Missouri) raised the issue during a visit to New Delhi.

Indian environmental groups have accused US companies of dumping surplus transgenic foods into aid packages, although India receives only a small amount of food aid every year. Indian researchers advising the government are against a blanket ban on genetically modified foods, but say that such foods should be labelled. India has licensed transgenic cotton, but has not sanctioned the commercial use of genetically modified foods.

Heavy fine for light-fingered archaeologist

San Diego An amateur archaeologist who removed skeletons and other artefacts from an American Indian site has been handed a civil fine of US\$2.5 million.

Officials allege that Jack Harelson of Grants Pass, Oregon, gained inside knowledge about the site — Elephant Mountain Cave in northwest Nevada — while working as a volunteer at the Nevada State Museum in Carson City during the 1980s. Harelson served 18 months in jail and paid a \$20,000 fine about five years ago for the crime, but the civil penal process was delayed by appeals until last month. The Department of Justice

says it will now move to seize Harelson's property and dock his wages.

Harelson maintains that he kept the skeletons and artefacts for a decade as part of an amateur research project, and says that he will appeal against the fine.

High-altitude observatory meets mounting opposition

Washington Plans for a facility designed to study star formation are facing some down-to-Earth problems. Native American tribes say that the Combined Array for Research in Millimeter-wave Astronomy, a radio telescope due to be built in the Inyo Mountains in southern California, will intrude on their ancestral lands. A local plant society has also voiced objections, claiming that the project will damage the region's pristine environment.

Astronomers argue that the array, which will be formed by merging two existing Californian radio telescopes, should be sited in the Inyo region, as its arid conditions and high altitude give ideal viewing conditions. The \$15-million facility, which would occupy about 300 hectares, is due to be completed by 2005. Leo Blitz, an astronomer at the University of California, Berkeley, says that the project's organizers will continue to work with local groups to resolve the problems.

Radio telescope plan gets inaugural chief

Amsterdam Ambitious plans to build the Square Kilometer Array, a radio telescope 100 times more powerful than any existing facility, are to be spearheaded by Richard Schilizzi, former director of the Joint Institute for Very Long Baseline Interferometry in Europe.

The increased power of the array will allow researchers to study very distant objects, making it possible to probe the era before stars and galaxies formed. Researchers will also be able to extend the search for intelligent life to millions more galaxies, and to reveal new pulsars, objects that emit pulses of radio waves.

Consortia in Australia, Canada, China,



Richard Schilizzi will lead the Square Kilometre Array project in its initial planning stages.

India, Europe and the United States are competing to host the array, which will cost about US\$1 billion to build. A design will be chosen in 2007, with the aim of beginning the five-year construction plan in 2010.

Schilizzi, currently at Leiden University in the Netherlands, was appointed as the first director of the project on 1 January.

► www.skatelescope.org

What do you get if you cross comedy with gene profiling?

Tokyo Blood samples, microarrays and two comedians — these are the ingredients for a study of the links between laughter and health.

Kazuo Murakami, a biochemist and professor emeritus of Tsukuba University in Japan, wants to determine whether laughter is good for you. On 12 January, comic duo B&B will perform in Tsukuba, and Murakami's team will examine what happens to the expression of genes linked to immunity and blood-sugar level in the audience. Both types of gene are thought to be affected by stress.

In one experiment, four Tsukuba University students will have blood taken before and after the performance. Researchers will use DNA microarrays to monitor the activity of 1,400 genes related to immunity. The blood sugar levels of 25 older audience members will also be tested. "People always talk about the health effects of negative stress. We want to study the effects of positive stress," says Murakami.

Double helix gets minty fresh image



London Fifty years after it became scientific currency, the structure of DNA is to be commemorated on a British coin. The Royal Mint is marking the anniversary of James Watson and Francis Crick's discovery by inscribing the famous double-helix configuration on a new £2 coin.

All £2 coins issued to banks by the Royal Mint for general circulation this year will feature the inscription. A separate commemorative coin made of silver and gold will also be produced, for sale to collectors and wealthy biologists.

Agriculture shock

Fears about terrorism usually centre on nuclear or biological weapons. But attackers could cause huge economic damage by spreading plant or animal diseases. Virginia Gewin asks how this threat is being confronted.

T. ARRIZA/CORBIS

The statistics on the 2001 outbreak of foot-and-mouth disease (FMD) in Britain make for disturbing reading. Four million cattle were culled to contain the disease, and estimates of the cost to the British economy — primarily to the agriculture and tourism industries — run as high as £30 billion (US\$48 billion).

The outbreak was unintentional, probably the result of illegal imports of infected meat being fed to pigs. But it can also be seen as an expensive warning. Although smallpox and anthrax receive the bulk of government and media attention when it comes to assessing the risk of bioterrorism, a deliberate attack on agriculture could disrupt trade and cripple agricultural industries. The risk of human fatalities or a serious food shortage is low, but few events would cause more economic damage than attacking the food supply. “The British FMD epidemic has given a blueprint to any terrorist,” says Martin Hugh-Jones, a veterinary epidemiologist at Louisiana State University in Baton Rouge.

If such an attack seems unlikely, it is worth noting that agricultural bioweapons have been used before. During the First World War, German agents infected Allied horses with the bacterium *Burkholderia mallei*, which causes glanders — a disease that can kill horses and can also infect humans. And Simon Whitby, a peace-studies researcher at the University of Bradford, UK, says that any country that has studied biological weapons, including the United States and Russia, will have looked at plant and animal diseases. Iraq, for example, is known to have weaponized wheat pathogens.

The main effects of any new attack are likely to be economic. The World Trade Organization lists few reasons for refusing to import crops and animals, but the presence of disease is one of them. Such bans can have rapid and severe consequences. When karnal bunt, a fungal disease of wheat, was found in northern Texas in 2001, over 25 countries banned wheat imports from the four infected counties within a single day. The estimated loss of revenue was \$27 million.

Agriculture is also now more open to attack, as a result of large-scale methods such



Close combat: could intensive factory farms become another battleground in the war on bioterrorism?

as the use of factory farms and monoculture cropping systems. "There is a vulnerability. It isn't anyone's fault, it's just how agriculture has evolved," says Jim Cook, a plant pathologist at Washington State University in Pullman. And although the ease with which a pathogen could be introduced is the root of the problem, weaknesses in the systems used to detect an outbreak could exacerbate any damage.

On the alert

Thankfully, the threat is now receiving attention. Funding in the United States has increased since the attacks on 11 September 2001 — President George W. Bush's proposed budget for the 2003 financial year includes an extra \$146 million to protect agriculture and the food supply, including money for monitoring animal health and for setting up a coordinated system to respond to disease outbreaks. And last September, the US National Academy of Sciences released *Countering Agricultural Bioterrorism*, a report detailing the problems that it says need to be addressed, from better diagnosis to improved communication between those who monitor potential outbreaks.

In Europe, defences are being boosted in response to an increase in the number of diseases that the continent's agriculture is expected to be exposed to, whether through increased trade or the possibility of climate change. "We have to be prepared to fight diseases that we didn't have to fight before," says Alex Thiermann of the Paris-based World Organization for Animal Health (OIE). Britain's FMD epidemic has also given a new urgency to research on diagnostic tools and vaccines, as well as encouraging countries to coordinate their plans for responding to future outbreaks.

Of all the lines of defence, rapid detection is considered most important. In many agricultural settings, surveillance systems usually rely on a farmer or local-government agent noticing something unusual. But the clinical symptoms of some diseases appear days after infection. With animals kept so closely confined, entire farms can rapidly become infected before anyone is aware of the problem. "Animals are kept in perfect environmental conditions for spread," says Larry Madden, a plant pathologist at Ohio State University in Columbus. It is not uncommon, for instance, to find 5,000 animals on 20 hectares of land in a typical US dairy operation. Rapid and extensive movement of animals between farms, slaughterhouses and markets also accentuates the problem. FMD, for example, had spread across Britain before it was detected.

The backbone of the global disease-detection system is the chain of 156 reference laboratories and collaborating centres run by the OIE, which analyse samples taken from animals that are thought to be infected. In the United States, the laboratory responsible for diagnosing exotic diseases is the Plum



Devastating diseases: accidental outbreaks of kernal bunt (above) and foot-and-mouth (left) have wreaked havoc on farming economies.

C. REDEL/AP

is short, and it is expected to undergo field trials in the next few months. "Those kits are extremely valuable if we have to make quick decisions," says Thiermann.

Another quick, but more sensitive, approach to FMD testing is being explored at Plum Island and Pirbright. Researchers have been evaluating field versions of a method for detecting viral RNA using the polymerase chain reaction (PCR), a common tool used to copy lengths of genetic material. Pirbright officials say they have a fully automated version of the test that could readily be established in a mobile lab should the need arise.

S. FRASER/SPL

Island Animal Disease Center in New York, run by the US Department of Agriculture (USDA). It currently conducts fewer than 1,000 tests every year for FMD, each of which is sent in to the lab. But an outbreak of FMD would require it to run tens of thousands of tests. And the quicker it could deliver the results, the sooner vets working in the field could decide what action to take.

'Pen-side' tests could help to speed up the process. Researchers at Britain's Institute for Animal Health in Pirbright, Surrey, for example, have adapted the laboratory procedure used to check for the presence of the FMD virus so that fieldworkers can perform the test by applying a treated paper to a tissue sample from a live animal. By monitoring the change in colour of the paper, the researcher can determine whether the antigens are present in about 10 minutes. The test has its drawbacks: false negatives can occur at rates of 10–20% when only small amounts of the FMD virus are present, and a cell-culture test is needed to confirm results that prove negative for the virus. But it could yet be a useful aid when time

Below the radar

Early detection of crop pathogens is also vital. The United States has more than four million square kilometres of farmland, much of it in remote areas where surveillance is virtually non-existent. Crop diseases can go unnoticed for a long time, during which they are continually spreading. It is estimated that the plum pox virus, discovered in Pennsylvania fields three years ago, was present for six to eight years before it was detected. In the past, a disease sample might have awaited identification until it eventually found its way to the often solitary plant clinician in the agricultural department of the local university.

The USDA is now attempting to improve this situation. Training modules are being developed for 'first responders', such as the farmers and crop consultants who are likely to be the first to notice a problem. And of the \$43 million allocated to agricultural-bioterrorism preparedness in the 2002 budget, \$20 million is earmarked to establish a network of diagnostic labs for plant and animal pathogens.



J. GIBB/AGSTOCK/SPL

Wide open: extensive monocultures such as maize can leave farms vulnerable to infections.

Five new regional centres are currently being set up at universities around the country, with the goal of providing rapid and accurate diagnosis of disease threats. Many existing labs will also get much-needed improvements. For example, the Great Plains Diagnostic Network at Kansas State University in Manhattan couldn't even run tests involving PCR until the recent funding arrived.

Once diagnoses have been made at the regional centres, all of the relevant information will be transferred to the National Agricultural Pest Information System, a database maintained at Purdue University in West Lafayette, Indiana.

To treat an infection once it has been detected, better knowledge of the pathogen involved is often required. Gaps in our understanding are being filled by new genomic sequences of animal and plant pathogens. In September 2002, for example, scientists at The Institute for Genomic Research (TIGR) in Rockville, Maryland, sequenced the genome of *Brucella suis* (V. G. DelVecchio *et al. Proc. Natl Acad. Sci. USA* **99**, 443–448; 2002), a pathogen that is considered a likely bioterrorism agent — the US military itself weaponized the bug in the 1950s. *B. suis* primarily affects animals, but can cause a debilitating disease in humans that can be lethal to people with weakened immune systems.

Surprisingly, the genome of this bacterium suggests that animal and plant pathogens are not as different as was once thought (see I. T. Paulsen *et al. Proc. Natl Acad. Sci. USA* **99**, 13148–13153; 2002). Many of the genes that control the metabolism of *B. suis* are also found in the plant pathogen *Agrobacterium*

tumefaciens, as well as in *Mesorhizobium loti*, a soil bacterium that forms a symbiotic relationship with plants. By investigating these links, TIGR researchers hope to reveal how *B. suis* survives outside its host, and thus generate new leads for tackling the pathogen.

Progress with plant diseases has been less impressive. As of September 2002, less than 6% of microbial genomes that had been sequenced and made publicly available belonged to plant-associated microbes. Animal pathogens are generally viruses or bacteria, but over three-quarters of plant pathogens are fungi, which have much larger genomes. "Compared to human and animal pathogens, plant pathogens are definitely behind," says Jacqueline Fletcher, president of the American Phytopathological Society.

Things are set to improve, albeit slowly. The US Department of Energy's Joint Genome Initiative announced in October that it will sequence two species of *Phytophthora*, a genus of fungus that is responsible for diseases as varied as sudden oak death syndrome and potato blight. What's more, a joint USDA and National Science Foundation programme has funded the sequencing of *Fusarium graminearum*, a fungus that causes disease in wheat and barley.

Forensic studies of deliberately caused outbreaks could also benefit from sequence data. Take the investigation of the US anthrax attacks, for example. Scientists had already sequenced the type of anthrax used — the Ames strain — and so were able to whittle down the number of possible places from which the bug could have been obtained. If sequences of different strains of other pathogens were available, investigators could eliminate many dead ends, as well as gaining incriminating evidence. "One of the differences is to be able to trace back and be able to get enough evidence to bring a case against the perpetrator," says Cook.

Take the strain

But realizing these ambitions will take hard work. Fourteen more strains of anthrax are being sequenced, and for other pathogens, many strains will also have to be sequenced if genomic databases are to be of any forensic use. Several researchers have called for the creation of a database of pathogen genomes, which could be used to investigate outbreaks of animal diseases, but it will take many years to gather enough information to respond to the range of possible bioterrorism agents.

More immediate improvements to our defences could come from new vaccines that are currently under development. Vaccines are an important means for dealing with animal disease, but existing versions are plagued with problems. FMD vaccines, for example, struggle to cope with different strains of the pathogen. "No single vaccine can bring immunity to more than a few of strains," says Mark

Wheelis, a bioterrorism expert at the University of California, Davis. Immunity is also often short-lived, and vaccinated animals cannot be distinguished from infected ones — and so are impossible to sell — because a weakened version of the live virus is used in the vaccine.

A new class of vaccines could counter some of these problems. Rather than using a version of the live virus, researchers are creating deleted, or 'subunit', vaccines, in which some genes have been removed. Proteins that are not crucial for antibody production, for example, can be removed from the vaccine. The treatment still prompts the production of the antibodies that protect against the virus, but vaccinated animals do not produce antibodies against the missing protein, and so can be distinguished from infected animals. Subunit vaccines have already been developed for Aujeszky's disease in pigs and bovine respiratory disease in cattle, and several groups are working on one for FMD.

Prepare for the worst

Like most recent developments, the vaccine work is driven by the need to contain an accidental outbreak of disease. But should countries be preparing specifically for a deliberate pathogen release? An intentional release could cause even more damage than an accident. The perpetrator could choose to release several highly virulent pathogens simultaneously in remote areas, for example. Authorities in the United States and Europe say that they are considering specific anti-terrorism measures, but many of the steps taken so far, such as the development of better diagnostic networks, tie in with conventional strategies for tackling plant and animal disease.

Given the scale of damage that a deliberate pathogen release could cause, some researchers say that stronger measures are needed. But others question this argument. In the bag of terrorist tricks, agricultural bioterrorism is the wild card, and some experts feel that it is unlikely to appeal to terrorists. "People attracted to terrorism wouldn't be as attracted to this," says Rocco Casagrande, who studies biological-agent detectors at Surface Logix in Brighton, Massachusetts. "Most terrorists are urban. They want a big bang. Killing cows or pigs is not a big bang," agrees Hugh-Jones.

But others warn that different approaches are likely to be used in future terrorist attacks. Whitby, for example, says that he and his colleagues consider an attack on agriculture to be the most likely form of biological terrorism that we're going to see. Predicting the behaviour of terrorists is, of course, notoriously difficult. But if economic damage is the aim, agricultural bioterrorism is certainly a viable threat. The size of that threat may be hard to define, but the financial scars left by Britain's FMD outbreak show just how serious it could be. ■

Virginia Gewin is a freelance writer in Corvallis, Oregon.

The oresmen

Some US organizations claim that fertilizing the oceans with iron could both help to tackle climate change, and make money. But marine researchers warn of unpredictable side effects. Quirin Schiermeier reports.

The *Ragland* does not look like a research vessel. The 100-year-old wooden Baltic schooner belongs to the Canadian rock star Neil Young who, between tours and recording sessions, relaxes on board the historic boat, which is usually anchored in San Francisco.

But last June, the *Ragland* was chartered for an unusual scientific mission. Young lent the boat to his Bay Area friend Russ George, a business consultant, amateur researcher and founder of the California-based Planktos Foundation. The foundation is one of several US organizations promoting what they say is an innovative method for tackling climate change, and George needed a boat to test his ideas.

After sailing to the Hawaiian Islands, George and his small crew dribbled a deep-red liquid — dissolved iron ore — into the sea. A few days later, the iron prompted a short-lived bloom of phytoplankton, single-celled algae that live at the ocean's surface. The algae absorb carbon dioxide, so fertilizing the sea to encourage their growth should increase the amount of CO₂ removed from the atmosphere, and hence help to tackle climate change. George and others reason that companies and governments that want to cut greenhouse-gas emissions will pay them to fertilize the oceans with iron.

As these groups begin to explore such ideas, ocean researchers are warning that the schemes could disrupt marine ecology and produce harmful gases. But despite their concerns, international law seems to be stacked against the dissenting researchers. As matters stand, there is little scientists can do to make companies perform thorough risk assessments on their fertilization plans.



Red trails in the sunset: Russ George (inset) hopes to tackle climate change using iron fertilizer at sea.

George's idea is workable in theory. Phytoplanktonic algae are responsible for about half of all of the biological absorption of CO₂. Most of the organisms pass through the marine food web and the CO₂ they have absorbed is returned to the atmosphere by respiration. But some will eventually sink to the ocean floor and remain there for hundreds of years, preventing the CO₂ from causing greenhouse warming.

Food for thought

The link between iron and phytoplankton was first highlighted by John Martin, formerly director of the Moss Landing Marine Laboratories in California. In 1988, Martin published a hypothesis suggesting that iron is the missing factor that limits phytoplankton growth in some ocean waters¹. The seas between Antarctica and the southern tips of the Americas and Africa, for example, contain plenty of some nutrients, such as nitrates and phosphates. But they don't contain much iron, or much phytoplankton.

Since Martin's death in 1993, several expeditions have shown that a lack of iron does indeed limit phytoplankton growth in these regions²⁻⁵. One such study — EISENEX — set sail from Cape Town in October 2000. More than 50 physicists, chemists and biologists travelled to the southern Atlantic on the German research vessel *Polarstern*, aiming to distribute iron in the eye of a 150-kilometre-wide eddy current.

Despite extreme weather conditions — the first gale hit the *Polarstern* five days after the iron solution was dissolved — the expedition was a scientific success. By monitoring the amount of phytoplankton in water samples taken before and after fertilization, the team backed up previous experiments that had shown that iron fertilization boosts phytoplankton growth. And by measuring phytoplankton concentrations at different depths, the researchers estimated how much carbon was transported to the deep ocean. They have yet to publish their results, but similar estimates by other teams suggest that

PLANKTOS FOUNDATION

the fraction of carbon transported is much lower than theory had predicted⁶.

Despite the uncertainty, these experiments have attracted the interest of entrepreneurs. The international effort to limit greenhouse-gas emissions, enshrined in the Kyoto Protocol, is based on a system of emission credits. Each country is assigned an emissions limit, but nations can exceed their target if they counteract their actions by financing measures that absorb carbon from the atmosphere, such as planting forests. Countries that have ratified the Kyoto Protocol are likely to impose emission limits on individual companies and run similar trading systems at the national level. And as projects such as planting new forests are costly, companies such as Planktos want to sell emission credits by running iron-fertilization projects. Although the United States has not ratified Kyoto, it may yet set its own mandatory emission targets, or choose to sign up to the protocol at a later date.

Ulf Riebesell, a marine biologist at the Alfred Wegener Institute for Polar and Marine Research in Bremerhaven, Germany, has compared studies of different methods for removing carbon dioxide from the atmosphere and says that iron fertilization would be 10–100 times cheaper than forestation, the next cheapest option. Riebesell also believes that large-scale, continuous fertilization of suitable areas of the oceans could allow between three billion and five billion tonnes of CO₂ to be drawn down per year — 10–20% of annual synthetic emissions.

Depth charges

No wonder, then, that the fight for pole position in a future ocean-engineering business has already begun. As well as the non-profit Planktos Foundation, which is financed by donations from energy companies and individuals, GreenSea Venture, based in Springfield, Virginia, says that it is planning large-scale commercial fertilization experiments. GreenSea's Michael Markels, a 76-year-old retired environmental engineer, has already obtained US patents on techniques for discharging iron fertilizers from floating buoys or commercial cargo ships.

"Until now, the number of field experiments has been restricted to those with major research budgets," says George. "By using an old sailing vessel, we were able to show that useful ocean research is not the exclusive domain of the richest researchers."

Although George's spirit of adventure may have caught the attention of the media, marine scientists are less enamoured by his ideas. They point out that there are no reliable

Sallie Chisholm fears that boosting algal growth may disrupt the oceans' ecology.



Steamed up: power generators may be interested in iron fertilization as a balance to their CO₂ emissions.

tools for verifying the amount of carbon taken up by the phytoplankton. More importantly, they worry about the impact of the fertilization on other marine life. "People like George and Markels claim that they can make the oceans green and solve all our problems, but it's not that easy," says Paul Falkowski, a marine biologist at Rutgers University in New Brunswick, New Jersey.

"There is no free lunch," agrees Sallie Chisholm, a marine biologist at the Massachusetts Institute of Technology who is at the forefront of attempts to resist the iron-fertilization plans. She fears that altering natural carbon fluxes could trigger a cascade of unwanted side effects. The iron could, for example, prompt the growth of toxic algae, which could kill other marine life or change water chemistry by removing oxygen. "The oceans are a tightly linked system, one part of which cannot be changed without it resonating through the whole system," says Chisholm.

Delicate balance

Riebesell adds that gaps in our knowledge about phytoplankton make it difficult to know which of the possible side effects will occur. "The life cycle of plankton and the evolutionary trends it has followed are completely different from those of terrestrial organisms," he says. "As long as our understanding of marine ecosystems is still in its infancy there is a great danger of abuse."

George counters such fears by pointing out that the input of iron into the ocean fluctuates with changes in the vegetation and farming activities on nearby land. Phytoplankton levels in the north Pacific have, for example, declined over the past 30 years. This seems to be linked to increased winter planting of wheat in China, which has limited the amount of iron-containing dust blown from fields into the sea⁷.

But there may be more at stake than marine ecology. Mark Lawrence, an atmospheric chemist at the Max Planck Institute for

Chemistry in Mainz, Germany, has looked at the possible disruption to climate that the fertilization could cause⁸. Phytoplanktonic algae produce dimethylsulphide, which influences cloud formation. The organisms are also believed to increase the amount of sunlight absorbed by the oceans. And they produce compounds such as methyl halides, which cause ozone depletion.

Lawrence doubts whether the developers of iron fertilization will voluntarily assess such possible environmental side effects. But currently there is no legal framework to demand a full environmental-impact assessment. International maritime law covers issues such as the dumping of waste material at sea, but contains nothing to prohibit commercial ocean fertilization.

And private groups such as the Planktos Foundation are not forced to use the same standards as academic researchers when experimenting on the open seas, points out John Cullen, a physical and biological oceanographer at Dalhousie University in Halifax, Canada. "Before any large-scale projects start, responsibilities, risks and possible benefits need to be addressed at an international level, with inclusion of all interested parties and stakeholders," says Cullen, who suggests that international bodies such as the United Nations Environment Programme should oversee all work on iron fertilization. But so far, politicians seem to have been deaf to such warnings, leaving organizations such as Planktos and GreenSea free to pursue their experiments in climate engineering.

Quirin Schiermeier is Nature's German correspondent.

1. Martin, J. H. & Fitzwater, S. E. *Nature* **331**, 341–343 (1988).
2. Martin, J. H. *et al.* *Nature* **371**, 123–129 (1994).
3. Coale, K. H. *et al.* *Nature* **383**, 495–501 (1996).
4. Boyd, P. W. *et al.* *Nature* **407**, 695–702 (2000).
5. Watson, A. J., Bakker, D. C. E., Ridgwell, A. J., Boyd, P. W. & Law, C. S. *Nature* **407**, 730–733 (2000).
6. Dalton, R. *Nature* **420**, 722 (2002).
7. Gregg, W. W. & Conkright, M. E. *Geophys. Res. Lett.* doi:10.1029/2002GL014689 (2002).
8. Lawrence, M. G. *Science* **297**, 1993 (2002).

Planktos Foundation
 ♦ www.planktos.com
 GreenSea Ventures
 ♦ www.greenseaventure.com

Funding should recognize the value of peer review

This service to science is threatened by time constraints and performance assessment.

Sir — The peer-review system, discussed in your News feature “Publish, and be damned...” (*Nature* **419**, 772–776; 2002), operates almost entirely on a voluntary basis. It carries some prestige, but does not really affect an individual’s academic progress. This is wrong, for two reasons.

First, to whichever end of the “scientific-publishing food chain” (as your feature engagingly calls it) it is applied, peer review plays a major role in preventing pollution of scientific data with falsified or distorted information. It also decreases wasteful publication of bad science.

Second, it introduces an element of additional, independent judgment, which eliminates noise and clarifies interpretation. Thus, like an audit in

the medical world, peer review is crucial in ensuring that science works properly.

In spite of all this, peer review is becoming more difficult to apply because of the increasingly scarce commodity of time. A substantial number of potential referees are declining to accept journals’ invitations to review, or are unable to deliver a report, because of time constraints.

Thus, there must be more institutional support of peer review. Although scientists and editors recognize that being invited to review is an act of academic recognition, assessment-conscious institutions fail to recognize this. Yet peer review is an important service to science, and it seems obvious that scientific institutions should support it very strongly. A ranking system

is needed, reflecting an individual’s participation in the peer-review process, similar to his or her publication record, and this should be taken into account in performance assessments. The system could be weighted according to the impact factors of the journals served, and there could be an element of funding associated with an institution’s peer-review score.

All this is necessary to take into account ever-increasing time pressures that could affect the quality of science. Peer review does not need to change fundamentally, but it should be moved higher up the list of academic priorities.

Marek H. Dominiczak

Medical Humanities Unit, Gartnavel General Hospital, 1053 Great Western Road, Glasgow G12 0YN, UK

Biology can be helpful to open-minded physicists

Sir — Your News Feature¹ “Bridging the culture gap” is a welcome update on the status of interdisciplinary research in biology, as relevant today as it was 15 years ago when Arthur Kornberg commented on a similar topic².

Communication is a major barrier that impedes collaboration between biologists and physicists.

Biologists tend to start by asking a specific question for a specific biological system. The answers to specific questions gradually contribute to the delineation of general principles.

Many physicists, when addressing biological problems, consider themselves as generalists who are more interested in developing general theories or finding evidence to fit them.

Although physicists can dismiss biologists’ explanations as too “empirical and descriptive”, this level of explanation is apparently good enough for biologists and has served biology well in practice. Physicists’ quantitative explanations may well provide a deeper level of mechanistic understanding, but deeper is not necessarily better if important biological context is lost. The satisfaction of scientific explanation is a relative term and exists in the eye of the beholder. Insisting on the most fundamental explanation for every biological phenomenon brings us all the way down to the elementary particles, which is beyond the scope not only of biology but also of many subdisciplines in physics.

The physicists who make the greatest strides in interdisciplinary research are open-minded, genuinely interested in biological problems and determined to make the transition. They apply quantitative techniques to address questions of interest to biologists.

John Hopfield is an excellent example of a physicist whose theoretical models have always focused on biological functions. In addition to his work on neural networks in the 1980s, as mentioned in your feature, in the 1970s Hopfield developed a series of elegant models on kinetic proofreading to explain the extremely low error rate in the biosynthesis of macromolecules and kinetic cooperativity of haemoglobin (see, for example, refs 3 and 4 respectively). These models and their variations have been enthusiastically embraced by biologists and are now being applied to areas of immunology, signal transduction, intracellular transport, DNA disentanglement and protein folding *in vivo* to explain why small quantitative changes in molecular interactions can lead to qualitative differences in biological function, as well as how biochemical energy can be used to improve the accuracy of molecular processes.

Although the idea of interdisciplinary research has been around for a long time and has been actively promoted by agencies such as the US National Science Foundation and Department of Energy, it is far from reaching its potential.

It is to be hoped that opportunities in genomics and other data-rich biological fields can finally provide the necessary attraction to drive more and more

interested physicists towards biology — and to keep them there.

Guangpu Li

Department of Biochemistry and Molecular Biology, University of Oklahoma Health Sciences Center, 940 S.L. Young Boulevard, BMSB 853, Oklahoma City, Oklahoma 73104, USA

1. Knight, J. *Nature* **419**, 244–246 (2002).
2. Kornberg, A. *Biochem.* **26**, 6888–6891 (1987).
3. Hopfield, J. J. *Proc. Natl Acad. Sci. USA* **77**, 5248–5252 (1980).
4. Hopfield, J. J., Shulman, R. G. & Ogawa, S. *J. Mol. Biol.* **61**, 425–433 (1971).

Strict guidelines make it clear who’s responsible

Sir — The headline “Physics guidelines drop equal-responsibility clause”, on your News story (*Nature* **420**, 258; 2002) reporting new ethical guidelines adopted by the American Physical Society (APS) in response to recent cases of misconduct, leaves the impression that the APS weakened its guidelines. On the contrary, the guidelines have been strengthened.

A major reason for the APS revision was to make explicit the responsibilities of co-authors. In particular, the new version includes a provision that specifies some authors who must take full responsibility for the content of an entire paper.

As your News story indicates, some physicists feel that the guidelines are now too strict, but most of them are supportive. The APS will continue to consider its role in establishing ethical guidelines and promoting education in professional conduct.

Myriam P. Sarachik

President, American Physical Society, 1 Physics Ellipse, College Park, Maryland 20740-3844, USA

When biodiversity meets humanity

Conservation may require political change to trust those who use the land.

Biodiversity, Sustainability and Human Communities: Protecting Beyond the Protected

edited by Tim O'Riordan & Susanne Stoll-Kleemann

Cambridge University Press: 2002. 334 pp.
£45, \$65 (hbk); £16.95, \$23 (pbk)

Brian Child

If you think about it, conservation is like the old Soviet Union. A few self-appointed wise men — scientists mostly, in this case — decide what is best for everyone, justified by a strong ideological creed of nature preservation, animal welfare and wilderness. With democratization encouraging the reintegration of humanity with biodiversity, the authors and editors of this book provide some useful ideas about how to do this. But they instinctively view reintegration through biological field-glasses, which are perhaps the wrong instrument.

Conservationists worldwide have not formulated clear concepts for sustainable resource use. They have failed to shift the focus from biological outcomes to political, institutional and economic causes, and to remove themselves from the locus of control. In the most lucid chapter, Norman Myers convincingly shows that the causes of biological loss are not biological but institutional; that global biodiversity is everyone's heritage but nobody's business; that poor people damage marginal environments that are rich in biodiversity; and that subsidizing habitat conversion (agriculture), extraction (forests, water, fish) and fossil fuels damages ecosystems. He begs conservationists to address these root causes, particularly the institutional roadblocks.

In its first six chapters and its final one, the book assesses the threats to biodiversity and the need to change paradigms and to include those people who are seeking solutions. Supporting material is supplied in seven chapters, which together present nine case studies. These include failures and tentative starts in 'new' countries (Estonia, Croatia and Indonesia), integrating people's aspirations with parks and landscape conservation (Cape Floral Kingdom, Europe and the United States), private involvement in Brazil and Costa Rica, and the commercialization of wildlife at landholder level in Namibia.

The book introduces concepts that promise better, more equitable resource allocation and governance — property rights, free choice and markets, access to justice and equitable local control — but does not build these into a guiding framework. There is an



Balance of power: children in the Solomon Islands failed to save this tree from a mining company.

emerging proprietorship—price—subsidiarity hypothesis, which says that if rights to natural resources are clearly delineated, if the value of wild commodities exceeds other land uses, and if people closest to the land capture most of this value, then renewable resources are likely to be husbanded better.

This hypothesis underpins the Namibian case study (by Markus Nuding), the book's only example of entrusting landholders with wild resources. The strategy shows promising results and presents a strong moral justification for landholder control. Passing resource control to millions of rural people at the periphery of power requires new systems of governance that maximize value while curbing overuse. Prohibitions, policemen and fines are increasingly unacceptable. Making the necessary shifts in power relations is the greatest challenge, however. Political and institutional economics are important tools for understanding this

transition, and conservationists need to arm themselves with these.

As our demands on the planet increase, our ability to govern and add value to biodiversity will determine how much is conserved and for how long, as well as its contribution to livelihoods. But we do not manage biological resources productively at present. 'Conservation' has been controlled, at national and international levels, by centralized, hegemonic institutions like those that have prevented societies in China, Arabia and, more recently, the Soviet Union from adapting to change. Central control arises from a fundamental mistrust of people and is managerially unproductive. Increasingly aware of their democratic rights, people in developing societies are challenging this old, inequitable (even imperialist) way of doing things.

Even in the inner sanctums of conservation — notably CITES and the IUCN's World Conservation Congresses — sustainable use

T.R. COOKE/UNEP

is gaining acceptance and trade bans are being overturned. The conventional conservation paradigm is being challenged because global interest cannot continue to impose costs and loss of benefits on local communities. The political economics of conservation is becoming critical — we are being forced to clarify who conservation is for, who will implement it, and how we will pay for it.

Responding to this challenge means turning conservation upside down, placing the right to benefit and the responsibility to husband squarely with landholders. This is a daunting prospect. It involves a choice between imposing conservation on society as a moral obligation, and believing in more democratic politics. This book's instinctive support of the former (with two exceptions, Nuding on Namibia and Ione Egler's strong and pragmatic assessment of Brazil) prevents it from fully exploring the latter.

Despite its ambitious title and the international rhetoric, the book reflects how little innovation is being allowed. Most of the nine case studies investigate the relationship between parks and the communities that they supposedly serve, albeit with a democratic twist, but 'beyond the protected' there is surprisingly little emphasis on making conservation an enterprise choice for landholders. Centralized conservation's mistrust of the private sector means that examples are too few and too constrained to be conclusive or to drive much innovation.

Yet conservation's failures, and growing resentment of imposed conservation, is driving change. Results from experimentation with people-centred conservation are mixed, often because initiatives are ill-conceived, over-cautious, or pay too little attention to institutional or economic principles. The book would have been strengthened by assessing why some initiatives worked and others did not. The danger is that the winners from centralized conservation — scientists at the global and national controls, and predatory élites who plunder forests, fish and wildlife — will use this inconclusiveness as an excuse not to trust people with resources. Because, messy as it is, only by entrusting rural people to manage wild resources in their own interests are we likely to conserve our planet.

This book makes a start in addressing these vital issues. Its strengths are its case studies and its important recognition that participatory democracy (summarized well, if academically) will, and must, drive a paradigm shift in conservation. Perhaps if it had taken this logic to its conclusion, and developed concepts more from the practicality of landholder motivation than the moral conviction and world-view of conservationists, even stronger hypotheses might have emerged. ■

Brian Child is chair of the Southern African Sustainable Use Specialist Group (IUCN), Lusaka, Zambia.

Aliens unlimited

Evolving the Alien: The Science of Extraterrestrial Life

by Jack Cohen & Ian Stewart

Ebury Press/John Wiley: 2002. 380 pp./288 pp.

£17.99/\$27.95

Lawrence M. Krauss

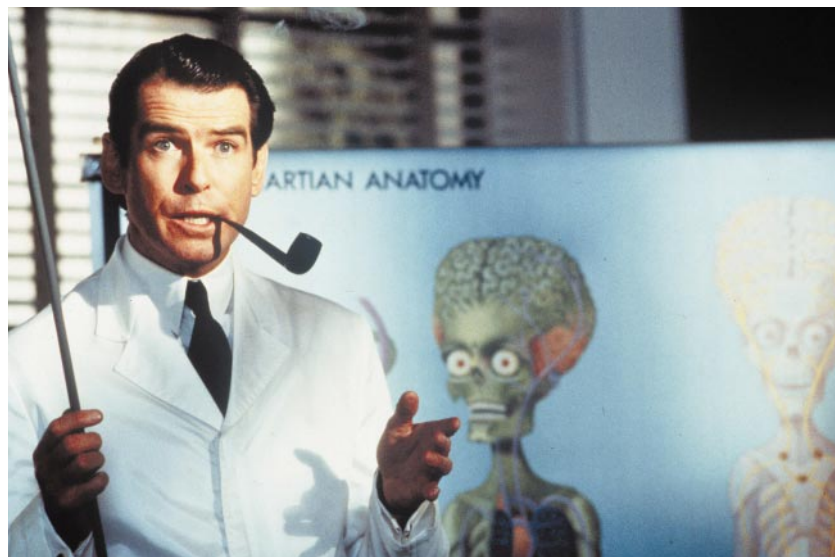
Jack Cohen and Ian Stewart are both well-established authors of popular science books, as well as being research scientists (in biology and mathematics, respectively) and aficionados of science fiction. Their new book focuses on the various possibilities of extraterrestrial life (normally associated with the emerging field of astrobiology), and on the relationship between science and science fiction. Both subjects have been heavily mined in the popular literature of late, so I admit that I was not initially enthusiastic about reviewing yet another book in these areas, but I felt that these authors' track records warranted at least a look.

The authors make it clear at the outset that the book is neither a homage to astrobiology, nor something like 'The Biology of (put your favourite science-fiction series here)'. Their arguments are: first, that astrobiology as they claim it is currently practised is largely irrelevant for imagining the spectrum of possibilities of extraterrestrial life; and second, that the biology of science-fiction aliens cannot be achieved by extrapolating science as we know it. I confess to being quite sympathetic to this latter claim. Indeed, in one of my favourite reviews of this genre of books, in *The Economist* several years ago, the reviewer lamented that although incorrect physics can often provide an interesting launching point for

speculation and wonder, incorrect biology is usually just silly. So I knew what Cohen and Stewart's new book is not, but I confess that I am still not completely clear what it is all about, or what the target audience might be.

I suspect that the authors intend the book as an antidote for the pessimism about alien life that was generated by Peter Ward and Donald Brownlee's widely reviewed *Rare Earth*, which argued that the conditions on Earth that were responsible for the formation of life are so special that life like that on Earth is likely to be rare at best in our Galaxy. Cohen and Stewart take the opposite view: that recent scientific discoveries vastly expand the possibilities for life elsewhere. The brunt of their wrath is directed against conventional astrobiology, which, they argue, is short-sighted in the extreme. They propose instead that most forms of life probably require conditions nothing like those found on Earth during its history.

Astrobiologists are not the only targets of their ridicule. Cohen and Stewart also find fault with, among others, evolutionary biologists, chemists, astrophysicists, Hollywood producers, most other science-fiction authors, and many classes of science-fiction reader (although I confess that they were quite kind to me). I think it is fine to have a bit of fun, but I found the use of these individuals as 'straw men' for their arguments relentless. Clearly, Cohen and Stewart dislike inaccurate scientific claims about extraterrestrial life, as well as the inaccurate science-fiction portrayals of such life. But people who are interested in the work of the groups that they target are likely to be the potential buyers of this book. Happily, the authors do a fine job in taking on the fallacies of creationists and alien-visitation fanatics, but I expect that here they are preaching to the choir.



Mars Attacks! But has science fiction been imaginative enough in guessing what aliens might look like?

Cohen and Stewart clearly enjoy science fiction at least as much as science, and the most amusing parts of the book involve sequences where they attempt to illustrate their points by inventing different science-fiction alien species. My favourite involves 'The Sensual Tribble'. They also show their clear affection for science fiction by incorporating throughout the text mini-summaries of many of the classic stories of this genre over the past 50 years. I found that this adds little to their argument and instead tends to be distracting. Alas, this may merely represent my own bias, which is that science fiction tends to be less imaginative, and therefore less interesting, than science, because nature generally exceeds the limits of the human imagination when it comes to exotic and unexpected phenomena.

Probably the most refreshing aspect of *Evolving the Alien* is the authors' continued insistence that when it comes to the biology of extraterrestrial life, which they dub 'xenoscience', the possibilities are probably literally limitless. The pair work hard to expand horizons, but at times I found that their optimism strains the bounds of credibility. When their scientific arguments stray away from biology, they sometimes lapse into proposals that seem to make no sense.

Arguments that aliens living during 'inflation' (which ostensibly occurred when the Universe was 10^{-35} seconds old) might somehow trick us about the current age of the Universe (10^{10} years), or that the vacuum of space-time might possess sufficient complexity to organize itself into some form of life by carrying out a complete thermodynamic work cycle, seem to me to be serious misconceptions of physics. The repeated assertion that the Sun is 10 billion years old, and that the Earth is between 5 billion and 6 billion years old, also grates, because we know with great accuracy that the Sun is about 4.5 billion years old and the Earth younger still.

My respect for Cohen and Stewart as popularizers of science meant that I wanted to enjoy this book. But ultimately, while I was relieved to find out what this book was not, I couldn't get a feel for what it actually is. I expect that ardent science-fiction enthusiasts who have the patience to work through the authors' polemics, and who enjoy the many references to the authors' favourite science-fiction works, may feel differently. At the very least, many readers of this book will be inspired to expand their horizons when pondering the remarkable possibilities for life in the Universe. ■

Lawrence M. Krauss is in the Department of Physics, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, Ohio 44106-7079, USA. He is the author of numerous popular-science books, including *The Physics of Star Trek* and *Atom: A Single Oxygen Atom's Odyssey from the Big Bang to Life on Earth ... and Beyond*.

A molecular view of the cell

Cell Biology

by T. Pollard & W. Earnshaw

Saunders: 2002. 834 pp. \$59.95/£34.99.

Electronic Image Collection (CD-ROM)

\$395/£425

Joe Howard

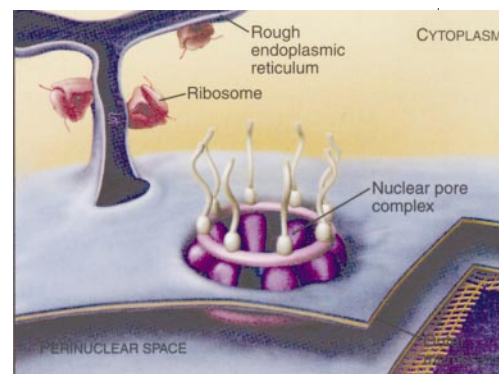
Do we need another cell-biology textbook? After all, the classic *Molecular Biology of the Cell* by Alberts *et al.* (Garland Science, 2002) is still going strong in its fourth edition; and *Molecular Cell Biology* by Lodish *et al.* (W. H. Freeman, 2000) and *The Cell: A Molecular Approach* by Cooper (Sinauer, 2000) are worthy alternatives. So there is no excuse for cell biologists, be they undergraduates or cutting-edge researchers, not to have a broad knowledge of their field. What, then, does this new textbook, *Cell Biology* by Pollard and Earnshaw, have to offer?

The first thing that struck me is that this is not a book about cells. Go to the index and you won't find any neurons, glia, chromaffin cells, keratinocytes, melanocytes, hepatocytes, myoblasts or hair cells. This is partly due to the poor indexing, but mainly it reflects the book's molecular emphasis. With a few exceptions, such as the cells of the blood and connective tissues, there is little detailed information about the biology of cells, and how cells are adapted for their functions in tissues and organisms. For this you'll have to go to a specialized text such as *Cell Movements* by Dennis Bray (Garland Science, 2000).

Unashamedly, *Cell Biology* is about molecules. As such it is a magnificent piece of work. Most of the chapters begin with the structures and family trees of the key molecules. In this post-genomic era, this is the logical way to organize our knowledge of biology. The danger is that the approach can be dry. But by focusing on mechanisms and principles, the book shows the connections between different cells, and between the different organisms. And this is gratifying.

Perhaps the most stunning feature of the book is its illustrations. Molecules leap out from every page. Proteins, DNA, membranes and small molecules are all beautifully rendered by Graham Johnson. Atomic structures are used when available. In each figure, molecules are drawn in proportion to create a vivid impression of the scale and intricacy of the cell's building blocks. And the book is lavishly illustrated with electron micrographs, many from Don Fawcett, one of the pioneers of cell biology. This goes a long way to redress the molecular bias. All of the illustrations are available on the accompanying CD-ROM.

Have the authors got their facts straight? For the most part, yes — this is a very scholarly text. But there are notable exceptions.



Cellular structures, such as the nuclear envelope, are important even in the post-genomic era.

For example, Figure 1.1A is a phylogenetic tree showing the radiation of the Eubacteria, Archaeobacteria and the Eucarya on the basis of rRNA-sequence comparisons. But the structure of this tree has been dramatically altered as a result of recent developments in molecular phylogeny. The problem is not that the information is out of date; biology is a living subject, so this is inevitable. Rather, this example illustrates how important it is to not present hypotheses as fact. Another example is the promulgation of the common misconception that the microtubule is the only cytoskeletal polymer that can resist compression. Yet the protrusion of moving cells is driven by the polymerization of actin filaments that must act in compression. One of my pet peeves is about units: there is no place for non-SI units such as the centimetre, and the use of the mole runs counter to the molecular spirit of the today's biology.

Now to the bottom line. First, *Cell Biology* is short, only half the length of Alberts *et al.* Second, it is based in bioinformatics and protein structure. And third, it contains a lot of data upon which knowledge in cell biology is based. The sections on the cytoskeleton and the cell cycle, Pollard and Earnshaw's research fields, are particularly strong in this respect. Thus *Cell Biology* is a higher-level textbook than *Molecular Biology of the Cell*. The illustrations and the inclusion of kinetics make it a superb choice for an advanced undergraduate or graduate textbook in cell biology. It is essential reading for all workers in the field. ■

Joe Howard is at the Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauerstrasse 108, 01307 Dresden, Germany.

More biology textbooks

Molecular Biology of the Cell, 4th edn. A Problems Approach

by John Wilson & Tim Hunt

Garland Science, £18.99, \$33.95

Study Companion to *Molecular Biology of the Cell*.

Molecular Principles of Animal Development

by Alfonso Martinez Arias & Alison Stewart

Oxford University Press, £28.99

Nature's insight into what makes scientists tick.

Summarize yourself in the form of a title of a paper in Nature.

Patterns of adaptive evolution and ecological diversification.

What was your first experiment as a child?

When I was eight, I mixed hydrochloric acid and ammonium hydroxide. I'm proud that I identified the product as 'sal ammoniac' — ammonium chloride. I also remember alarming my parents with the odour.

Who has been the most important mentor in your career?

My PhD adviser, Sue Wessler, who gave me the freedom to do some of the first molecular evolution work in her laboratory. That laid the foundation for my current scientific interests.

What makes a good scientific mentor?

Someone who is both supportive and challenging, and makes sure you don't get away with sloppy thinking.

What single scientific paper or talk changed your career path?

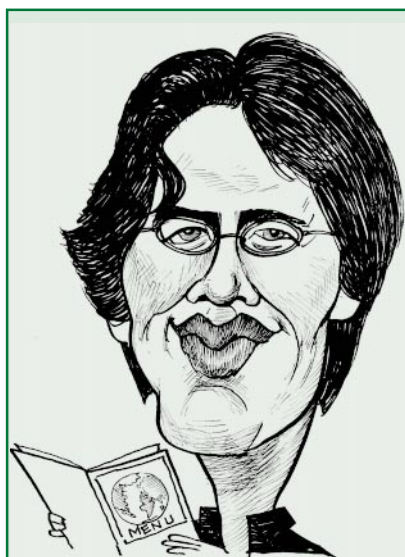
'Adaptive evolution in the stomach lysozymes of foregut fermenters' by Caro-Beth Stewart (C. B. Stewart, J. W. Schilling and A. C. Wilson *Nature* **330**, 401–404; 1987). I read it while browsing in the library at Columbia where I was a graduate student. It got me thinking about the molecular basis for evolutionary adaptations.

What book has been most influential in your scientific career?

One day I went into a used bookstore across the street from the café where I was studying and I saw William Bateson's *Problems of Genetics*. I started to read and was engrossed. Two things struck me: first, that the early geneticists were describing phenomena that we could more readily interpret today (I scribbled notes like "transposable elements!" on the margins). The second was that early geneticists were looking at species that would never be considered 'model organisms' today. This book made me begin to study the evolutionary genetics of natural populations using the modern tools of molecular and developmental genetics.

What literary character would you employ as a postdoc?

P. G. Wodehouse's Jeeves. He is highly intelligent (it's all that fish he eats), able to solve complex problems, unflappable, has a mastery of detail and is well-read. My current postdocs, however, although they have many of these qualities, lack one Jeeves trait — what Bertie Wooster describes as 'the feudal spirit'. It may be for the best.



Michael Purugganan

Michael Purugganan is an associate professor of genetics at North Carolina State University in Raleigh, where he studies plant molecular evolution. He likes modern art, tequila and is inordinately fond of black clothes.

What's your favourite conference destination, and why?

Bodega Bay, California. There is nothing to do there but eat barbecued oysters and sample the local vintages.

What book currently resides on your bedside table?

Felipe Fernández-Armesto's *Civilizations*, and a collection of Dorothy L. Sayers' detective stories featuring her *bon-vivant* sleuth Lord Peter Wimsey.

What music heads the playlist in your car or laboratory?

The Joshua Tree by U2, and a compilation of the Verve recordings of the Brazilian bossa-nova master Antonio Carlos Jobim. The latter keeps me calm when stuck in traffic. Road rage is such an ugly thing.

Where and when would you most like to have lived or worked?

I would love to live 1,000 years from now and see how it all works out.

Assuming the dead can be raised and/or time travel exists, who from the world outside science would you most like to have dinner with?

Paul Gauguin, Gabriel Garcia Marquez — and José Rizal, the nineteenth-century Filipino nationalist writer, scholar and

renaissance man whose work helped spark the Philippine revolution against Spain.

You are on a plane behind two students obviously going to the same conference, who start to talk about your work. What do you do?

I would join in without saying who I am.

What's the best piece of advice you've ever received?

Never, ever take anything for granted.

What do you most dislike about having research published?

People treat publications as if they are carved in stone. Research is a dynamic enterprise, and a publication is only a chapter in ongoing debates.

The Internet is the bane of scientists' lives because...

I never go to the library any more. I used to go at least once a week and browse the journals. It was a great way to keep up, and glance at papers far from my own interests.

What do you do to relax?

I'm a big foodie, so I'm always on the lookout for good restaurants. Some of my lab folks think that if they give me the name of a city I can probably tell them where to get a great meal.

What would you have become, if not a scientist?

I moonlighted as a journalist to put myself through college in the Philippines, writing on politics and economics — not the safest thing to do in a dictatorship. When I was in my third year at the University of the Philippines, I got a call from the Manila bureau chief of Associated Press (AP) and offered a position — on the spot — as one of AP's Manila correspondents. I turned it down, but it was tempting.

What single discovery, invention or innovation would most improve your life?

Do you think matter–energy transport is possible?

Name one extravagance you can now get away with because of your eminence.

Not having to buy the cheap airline ticket with the Saturday-night stayover.

What music would you have played at your funeral?

John Williams' *The Raiders March*, the theme for the Indiana Jones movies. Death is just another adventure.

What's just around the corner?

Evolutionary and ecological genomics. It's already happening, and suggests the ability to scan genomes and study links between molecular, organismal and ecological phenomena. ■

Act natural

Steven A. Benner

Over the past two centuries, organic chemists developed their discipline by purifying natural products from biological systems and determining how they are assembled from their constituent atoms. As their deconstructive powers grew, chemists targeted larger biomolecules, first proteins and nucleic acids, and then supramolecular structures such as the nucleosome and the ribosome.

Sequencing the human genome represents a culmination, of sorts, of this type of chemical reductionism. After all, the human genome is nothing more (and nothing less) than a statement about how carbon, hydrogen, nitrogen, oxygen and phosphorus atoms are bonded together in the organic molecules that we inherit.

Traditionally, however, chemists complemented deconstruction with a second theme: synthesis. After determining the structure of a new natural product, organic chemists made some more of it. At first, chemists used synthesis simply to prove that the structure proposed for their natural product was correct. Over time, however, synthesis became a struggle that pits the wits of chemists against increasingly complex natural products. The struggle forced chemists to discover things about molecules. Perhaps best known are the observations made during the total synthesis of vitamin B12 that led to the Woodward–Hoffman rules of orbital symmetry, recognized by the Nobel Prize in 1981.

Chemists next attempted to synthesize molecular structures that were not identical to natural products, but reproduced some of the behaviours that natural biomolecules display. 'Biomimetic chemistry', as this endeavour came to be known, was more than the preparation of the plastics, pharmaceuticals and other synthetic materials of modern civilization. The goal became to reproduce advanced.

dynamic behaviours of biological systems, including genetics, inheritance and evolution. These goals define synthetic biology as a field.

To a synthetic biologist, life is a special kind of chemistry, one that combines a frequently encountered property of organic molecules (the ability to undergo spontaneous transformation) with an uncommon property (the ability to direct the synthesis of self-copies), in a way that allows transformed molecular structures themselves to be copied. Any chemical system that combines these properties will be able to undergo darwinian selection, evolving in structure to replicate more efficiently. In a word, 'life' will have been created.

But what chemical structures combine these properties? Computer models that simulate replication and evolution *in silico* are relatively easy to come by. A computer program can suffer mutations and keep on ticking. But real molecules often change their behaviour dramatically upon even a slight change in structure. Chemists have in hand a modest number of chemical systems that can function as templates for their own synthesis. But those that can suffer mutation and still have 'children' are proving harder to find.

The pursuit of synthetic biology has already identified some structural features of natural self-replicating biomolecules that are crucial to their performance. For example, the repeating negative charges on DNA's phosphate backbone allow it to survive mutation while continuing to act as a template for copies. These charges are more important for evolution than the identity of the specific bases that carry the genetic information.

From such observations have come entirely new, artificial genetic systems. The first, created a decade ago, contains 12 bases rather than the four (A, T, G and C) found in natural DNA. The artificial genetic system can direct the synthesis of artificial RNA and proteins with extra amino acids. Chemists have appended functional groups onto synthetic genetic systems, creating a hybrid biomolecule that can be copied like DNA, but which also carries the functionality of proteins.

In parallel, molecular biologists have recruited tools from natural systems to generate synthetic evolution in the test tube. To identify new receptors and catalysts, libraries of random nucleic-acid sequences are generated and challenged in an artificial environment. Those that perform well are allowed to direct the synthesis of their copies. Those that fail are discarded. Chemists and molecular

Synthetic biology

This burgeoning field aims to reproduce advanced, dynamic behaviours of biological systems, including genetics, inheritance and evolution.

biologists are now set to merge the two, creating artificial evolution with artificial genetic systems from synthetic biology.

Synthetic biology is not merely demiurgy. In September, for example, the US Food and Drug Administration approved a diagnostic test that uses artificial genetics to measure the level of the HIV virus in infected individuals. The artificial genetic system allows high sensitivity and remarkable dynamic range, without interference from the natural DNA found in patient samples, and represents a remarkable step forward in the management of AIDS as a disease.

Synthetic biology should also help NASA to seek life in its probes of the Solar System. By asking what is possible in the chemistry that supports life, we are more likely to recognize weird life should we encounter it.

Perhaps most important, however, is the vista that synthetic biology offers at the frontier between chemistry and biology. Biology has been limited by “poke and measure the response” research strategies. Recombinant DNA technology allowed biologists to take the next step through synthesis of different sequences of natural molecules.

But synthetic biology with ‘bottom-up’ design may achieve much more. Building artificial genetic, regulatory and metabolic systems should offer a new way to learn more about genetic, regulatory and metabolic systems in general (and, dare we say it, universally). Thus, in this century, synthetic biology should aid the discovery of new, universal ideas about biology, ideas that might have remained undiscovered using simply reductionist analyses. Which is what synthesis did for chemistry in the twentieth century. ■

Steven A. Benner is in the Departments of Chemistry, and Anatomy and Cell Biology, University of Florida, Gainesville, Florida 32611, USA.

FURTHER READING

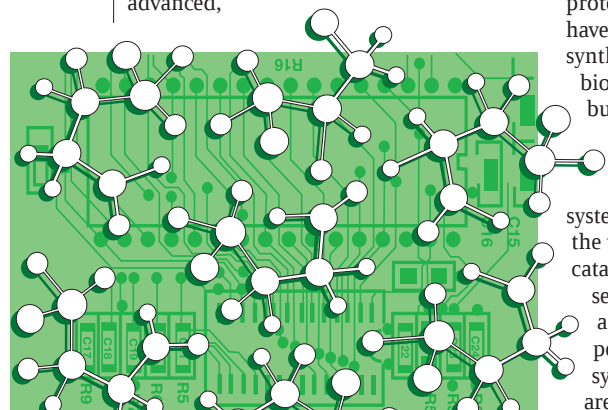
Woodward, R. B. *Pure Appl. Chem.* **17**, 519–547 (1968).
<http://www.genarts.com/karl/evolved-virtual-creatures.html>

Bain, J. D., Chamberlin, A. R., Switzer, C. Y. & Benner, S. A. *Nature* **356**, 537–539 (1992).

<http://www.alife.org/index.php>

Gee, H. *Nature* **409**, 1079 (2001).

<http://xanadu.mgh.harvard.edu/szostakweb/web.2>



Model for magnetic mystery

Maria T. Zuber

Today, the Moon has no magnetic field, but analyses of lunar rocks suggest that it did in the past. Did changes in the lunar interior create a magnetic dynamo billions of years ago?

Magnetic fields provide key information in reconstructing the evolutionary history of planetary bodies. Planets such as Earth and the gas giants have magnetic fields produced by self-sustaining dynamos that are a manifestation of the cooling of the deep planetary interior. In contrast, mystery surrounds the possible existence of a dynamo on the Moon. At present, the Moon has no internally generated magnetic field, but rock samples brought back by the Apollo missions bear evidence that the surface was magnetized in the past. On page 143 of this issue, Stegman and colleagues¹ revisit the question of dynamo generation on the Moon, making use of three-dimensional models that simulate the dynamics of the early lunar mantle.

On the surfaces of solid planets, some crustal rocks preserve the record of magnetic fields during earlier epochs of planetary history. The process of 'thermoremanent magnetization' occurs when certain volcanic rocks cool in the presence of a magnetic field. When the temperature falls past a value known as the Curie point, domains within magnetically susceptible rocks align and information on orientation and magnitude of the field at the time of cooling is preserved. For example, Earth's sea floor and the surface of Mars contain rocks that were magnetized in the presence of global magnetic fields generated by core dynamos, within the past 200 million years for Earth and over 4 billion years ago for Mars.

Although the Moon currently does not have an internally generated magnetic field, the Apollo samples show remanent magnetization that dates to between 3.6 billion and 3.9 billion years ago, half a billion to a billion years after the Moon formed. This is also the time of formation of the mare basalts — massive flood lavas that fill large impact basins on the near side of the Moon (Fig. 1). There has been much debate as to whether the lunar rocks were magnetized in the presence of a core dynamo because simulations have suggested that the interior would not have cooled rapidly enough to drive a dynamo².

The basic ingredients for dynamo action are an electrically conducting fluid within the planet (such as a liquid iron core), an energy source (such as convection arising from core cooling) and rotation (to organize the fluid motion). Various factors conspire against the idea of a vigorously convecting

Figure 1 The Moon's magnetic past. This image of the Moon, taken by the Galileo spacecraft en route to Jupiter, shows the bright highland areas, rich in anorthosite, and dark 'mare basalt' flood lavas. Stegman *et al.*¹ tie these features in to a model for how the Moon might have briefly generated a magnetic field 4 billion years ago.



NASA

primordial Moon, the most important of which are a small core and an 'inverted' thermal structure in the lunar interior. With regard to the latter, planets tend to be hot on the inside and cold on the outside, with internal heat being removed by convection. But analyses of Moon rocks have shown that the early Moon had a molten exterior. The pervasive presence of the rock anorthosite in the lunar highlands is evidence of this 'magma ocean' (Fig. 1).

A key aspect of the model of Stegman *et al.*¹ is that it takes into account the thermal and chemical consequences of the magma ocean. In a melting material that has the likely composition of the lunar mantle, minerals containing the lighter elements would float to the surface, forming the anorthosite crust. Left behind is a region enriched in heavy elements that is denser than its surroundings — so much so that previous simulations³ have shown it could sink to the deep lunar interior, possibly encircling the core. This layer would be rich in radioactive elements, such as uranium and thorium, which could produce sufficient heat as they decay to heat the Moon's interior significantly on timescales of hundreds of millions of years. In the model of Stegman *et al.*, the sunken residual material encircles the lunar core and insulates it from the rest of the mantle, trapping heat in the core and preventing the core from

cooling convectively, and also from developing a dynamo. But, after a period of time, the radiogenic material within this 'thermal blanket' decays and heats up. Eventually the material becomes more buoyant than its surroundings and rises towards the surface. With the removal of the thermal blanket, the core is then able to convect vigorously to cool itself, and this can produce a short-lived dynamo.

An essential element of the model is the timing. The length of time that it takes the thermal blanket to heat up and rise back towards the surface is broadly consistent with two important events in early lunar history — the eruption of the mare basalts onto the lunar surface and the magnetization of lunar rocks. Even so, the model is not without shortcomings. The early thermal and chemical structure of the Moon, which forms boundary conditions for the model, is highly uncertain. Although the core may be free to convect for the conditions assumed, the process may not be energetic enough to explain the magnitude of magnetization of lunar rocks. Even by adding additional energy associated with 'freezing out' the liquid iron core, the model can just barely explain the palaeointensity of recently reanalysed lunar samples (C. Johnson, personal communication).

Some researchers continue to doubt that an internal dynamo is required to explain the

magnetization of the Moon, and favour an alternative idea — that the observed magnetic signature was generated in association with large impacts during early history. In this model, magnetization should be concentrated diametrically opposite to major impact basins. Satellite observations from the Lunar Prospector mission show concentrations of crustal magnetization at the antipodes of some but not all large impact structures⁴.

Although the results of the current study are not proof of an early lunar dynamo, they do demonstrate that it is theoretically possible for a dynamo to explain both the timing of the observed magnetization of the lunar crust and the eruption of mare basalts. Fur-

ther simulations, continued analysis of the palaeomagnetic properties of lunar samples and low-altitude global mapping of the Moon's magnetic signature will be required for a true grasp of the elusive nature of lunar magnetism. ■

Maria T. Zuber, currently at Harvard University, is in the Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. e-mail: zuber@mit.edu

1. Stegman, D. R., Jellinek, A. M., Zatman, S. A., Baumgardner, J. R. & Richards, M. A. *Nature* **421**, 143–146 (2003).
2. Stevenson, D. J. *Icarus* **54**, 466–489 (1983).
3. Hess, P. C. & Parmentier, E. M. *Earth Planet. Sci. Lett.* **134**, 501–514 (1995).
4. <http://lunar.arc.nasa.gov/results/magelres.htm>

Developmental biology

A larval revelation

Thurston Lacalli

Identification of the previously unknown larval forms of the sea lilies, a group of marine invertebrates, is a refreshing reminder of the value of descriptive science in evolutionary studies.

The stalked crinoids, or sea lilies, are generally accepted as the most ancient of living echinoderms, with a fossil record extending back some 500 million years, almost to the base of the Palaeozoic^{1,2}. They live in deep water, and are difficult to collect in good enough condition for laboratory study. In consequence, their embryos and larvae have never been observed — until now, that is. On page 158 of this issue³, Nakano *et al.* describe how they took advantage of the local crinoid populations in two deep bays, southwest of Tokyo, where adult animals come to within 100 m of the surface. Using gill nets rather than dredges and trawls, and enlisting the help of local fishermen, the authors obtained healthy, gravid individuals that spawned in the lab and produced viable embryos and larvae.

The main significance of this discovery lies in the rich evolutionary connections of the phylum Echinodermata and the information that such 'missing' larval forms contain. The larvae of marine invertebrates are delightfully diverse morphologically, and often delicately beautiful. Most are small (less than 1 mm high) and use bands of cilia, arranged in diverse ways, for feeding and locomotion. There is frequently a 'type' larva for a given phylum, and similarities between larval forms can reveal evolutionary links between phyla that are not always evident from the morphology of the adult (Fig. 1). Zoologists with an evolutionary bent are thus keenly interested in any report of a previously undescribed larva, especially if it belongs to an organism as evidently ancient as the sea lilies.

Nakano *et al.*³ find that the development

of sea lilies resembles that of feather stars, their better-known shallow-water cousins, in having a stage known as a doliolaria (Fig. 2). The doliolaria is a yolky and rather featureless ovoid, with a regular series of four or five hoop-like ciliary bands, that settles without feeding and metamorphoses into the juvenile crinoid. In sea lilies, however, the doliolaria is preceded by a stage that has sinuous, longitudinally oriented ciliary bands. Although this stage also does not feed, it resembles the feeding larvae of other echinoderms — sea stars, sea urchins, brittle stars and sea cucumbers — which have ciliary bands arranged in sinuous patterns along the sides of the larval body. Of these, the newly discovered larva is most similar to the auricularia larva of sea cucumbers, although it is somewhat simpler in shape. This is consistent with its having secondarily lost the

ability to feed during evolution, something that has occurred repeatedly in other echinoderm lineages⁴.

The possession of a ciliated plankton-feeding larva has generally been thought to be a primitive feature of echinoderms, in part because a closely related group of marine invertebrates, the hemichordates, have a similar stage. It has always been something of an embarrassment, therefore, that no such larva was known from crinoids, ostensibly the most primitive group. The new results support the contention that the ancestral life history of echinoderms involved both an auricularia-like feeding stage and a non-feeding doliolaria. This is of broader significance for two reasons.

As 'basal' members of the Deuterostomia, the group of phyla to which vertebrates belong, echinoderms are a potential source of clues to the origin of chordates (the phylum that includes vertebrates). This is a major evolutionary puzzle, and several of the more reasonable explanations suggest that chordates are derived from larvae that were very like the auricularia. In the best known of these hypotheses, formulated by Walter Garstang in 1894, the characteristic dorsal nerve cord of chordates is generated by folding together the bilaterally paired, longitudinally oriented, ciliary bands of just such a larva. The idea that ciliary bands might convert directly into nerve cord in this way is now less widely accepted, but ciliary bands are the chief sites of nerve formation in auricularia-type larvae, so their involvement in the evolution of the chordate nervous system is nevertheless still a reasonable supposition.

The arrangement of the bands in ancestral larvae is thus a consideration in any serious attempt to reconstruct the events leading to the chordate condition. Circumferential doliolaria-type arrangements need to be considered as well as longitudinal ones. Converting the former into a nerve cord would require some rearrangement, but echinoderm larvae are accomplished at manipulating the patterning of epithelia and ciliary bands during development⁵. The evolutionary significance of this ability could be considerable but is not yet at all understood.

The second issue concerns the survival pattern of larval characters during evolution. If the possession of both an auricularia-like feeding stage and a doliolaria is ancestral, then among modern echinoderms the sea cucumbers preserve the primitive pattern most fully. Yet molecular phylogenies and morphological analyses based on adult characters routinely place sea cucumbers near the top of phylogenetic trees, as a modified (derived) 'crown' group^{6,7}. Determining whether morphological characters are primitive or derived has always been difficult. But the sea cucumber is a clear case of an organism that is considered advanced and



Figure 1 Picture of lily — the stalked crinoid *Metacrinus rotundus*. (Photo courtesy of S. Amemiya.)

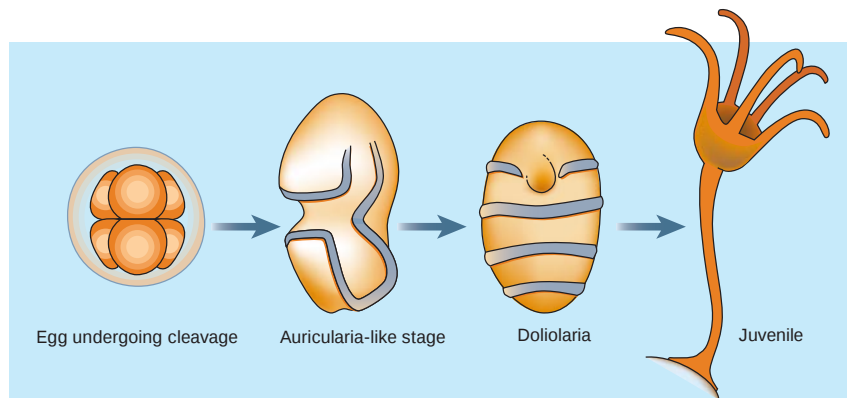


Figure 2 Stages in stalked crinoid (sea lily) development, as discovered by Nakano *et al.*³. Embryonic cleavage gives rise to an auricularia-like stage, which is non-feeding, as also occurs in some sea cucumbers. Ciliary bands are shown in blue. Transformation to a doliolaria follows, then settlement and development of the juvenile crinoid on the sea floor. Signs of sinuous ciliary bands have been observed in the related feather stars, but this is the first evidence that the doliolaria stage of any crinoid is preceded by a convincingly auricularia-like form. In life, the sea-lily larval forms are 0.5–0.6 mm high, and the stalked juveniles probably 2–3 mm when they begin to feed.

somewhat derived by most criteria, yet has a primitive developmental pattern.

The same problem crops up elsewhere among invertebrates, especially in the more primitive groups. The nemertine worms are an example. They are basal protostomes (the other major grouping of animals besides deuterostomes), but their larva is considered to be derived and of little phylogenetic interest because it occurs only in the most advanced members of the phylum. The echinoderm example suggests that this type of reasoning may be fundamentally flawed — that is, that larvae need to be assessed on their own merits, and that phylogenies based on pooled characters (and particularly adult ones) may seriously mislead anyone attempt-

ing to deduce the nature and life history of the ancestral members of a given group. In short, there is still a degree of art to the latter exercise that requires keeping an open mind about even well-accepted phylogenies.

Thurston Lacalli is at the Department of Biology, University of Victoria, Victoria, British Columbia V8W 3N5, Canada.

e-mail: lacalli@uvic.ca

1. Paul, C. R. C. & Smith, A. B. *Biol. Rev.* **59**, 443–481 (1984).
2. Mooi, R. *Can. J. Zool.* **79**, 1209–1231 (2001).
3. Nakano, H., Hibino, T., Oji, T., Hara, Y. & Amemiya, S. *Nature* **421**, 158–160 (2003).
4. McEdward, L. R. & Miner, B. G. *Can. J. Zool.* **79**, 1125–1170 (2001).
5. Lacalli, T. C. & West, J. E. *Invert. Biol.* **119**, 421–432 (2000).
6. Wada, H. & Satoh, N. *J. Mol. Evol.* **38**, 41–49 (1994).
7. Littlewood, D. T. J., Smith, A. B., Clough, K. A. & Emson, R. H. *Biol. J. Linn. Soc.* **61**, 409–437 (1997).

Global change

Leads, lags and the tropics

Robert B. Dunbar

A method that circumvents the problems of correlating different data sets has allowed the sequence of events at the last great deglaciation to be seen in finer detail.

Assessing how events in the tropics drive or amplify climate change has become a central issue for understanding climate in the past and predicting its future course. But there has been all too little information to go on. The instrumental record from these, the warmest parts of our planet, is woefully short. And although we have a reasonable appreciation of short-term phenomena such as El Niño, our knowledge of events and their effects on timescales of hundreds to thousands of years is much more limited.

One approach, as taken by Visser *et al.* (page 152 of this issue¹), is to determine the leads and lags between climate change at the

poles and at lower latitudes. The authors employed an exciting new strategy, using paired chemical and isotopic signatures in marine sediments from the Indonesian archipelago, to show that a rise in ocean temperatures in this part of the tropics played a key role in amplifying an episode of past global warming. Their sample site is in the part of the tropics known as the 'Indo-Pacific warm pool'. This is the biggest warm-water region on Earth, and so of all tropical regions this one is especially likely to have a large influence on climate.

Before the twentieth century, the last large and truly global change in climate occurred as a series of warming steps and

cold reversals between 20,000 and 10,000 years ago. This was at the end of the last ice age when, starting about 18,000 years ago, enormous ice sheets, mostly in North America and Scandinavia, melted away and produced a global rise in sea level of 120 metres. The amount of CO₂ in the atmosphere rose by nearly 90 parts per million (p.p.m.), from a low of about 180 p.p.m. in glacial times² (for context, 90 p.p.m. is roughly the same increase as that observed in the past 140 years, in this case from a starting point of 270 p.p.m.). At the same time, parts of the Earth warmed up by 10–12 °C.

The last ice age was but one of many, and we now know that subtle changes in the Earth's orbital characteristics serve collectively as a kind of climatic metronome, pacing the cyclical advance and retreat of continental ice sheets^{3,4}. Yet several mysteries remain. Chief among them is discovering how such small changes in the orbitally determined distribution of solar radiation are amplified to cause such large shifts in global climate and ice volume. Most explanations invoke a feedback loop caused by various mechanisms: movement of glaciers or retreat of sea ice, with attendant changes in their reflectivity; alteration of the ocean transport of heat; and changes in the concentration of greenhouse gases, such as water vapour and CO₂, in the atmosphere. One seemingly simple strategy for identifying the cause of amplification is to determine the order in which things happened during deglaciation. For example, if the great ice sheets began to melt and temperatures rose before there were large changes in the amount of CO₂ in the atmosphere, then we might conclude that CO₂ levels are not as important as other factors in the triggering or initial amplification of large-scale climatic change.

Analyses of ice and sediment cores around the globe have provided insight into climate change at hundreds of locations. Yet the goal of elucidating leads and lags during deglaciation has remained elusive, because it is surprisingly difficult to put all of the necessary time series onto a common timescale. Polar ice cores are sometimes dated by counting annual layers, and data from different cores can often be correlated using the isotopic chemistry of their constituent water or trapped gases. On the other hand, marine sediments are often dated using radiocarbon or other radionuclides, and there is an inevitable level of uncertainty about the age of the water in which the dating target formed, as well as the exact calibration to calendar years. The result is that the literature dealing with changes in ocean temperatures, the terrestrial records of ice-sheet retreat, and the ice-core evidence for changing CO₂ levels and polar temperatures, is full of seeming contradictions.

Visser *et al.*¹ have taken an elegant

approach that avoids the problem of using different time series to track different phenomena. They exploit a relatively new method of measuring past changes in ocean temperature^{5,6}, the Mg/Ca ratio of the carbonate shells of surface-dwelling microplankton called foraminifera. Armed with an Mg/Ca-based estimate of changes in sea surface temperature within the Indonesian archipelago during deglaciation, they then use a venerable isotopic tracer — the ratio of $^{18}\text{O}/^{16}\text{O}$ in these same foraminiferal shells — to determine the changing volume of ice in continental glaciers. This is possible because the oxygen isotopic ratio of foraminifera records both seawater temperature and its isotopic composition (largely controlled by the growth or melting of ice sheets, which are depleted in ^{18}O compared with sea water). An independent temperature estimate means that the foraminiferal isotopic ratios can be corrected for temperature effects, leaving a residual that is directly related to the amount of polar ice that has melted into the ocean during deglaciation.

Visser *et al.* present two important findings. They conclude that, 18,000 years ago, sea surface temperatures in the Indo-Pacific warm pool were 3.5–4 °C cooler than they are now — this is yet another piece of evidence that contradicts previous views that tropical temperature changes during glacial–interglacial cycles were much smaller, only a degree or two, than that. More notably, their work provides strong evidence that the last deglaciation was ‘felt’ first in the tropics, as shown by a rise in temperature within the Indo-Pacific warm pool, and that this warming signal propagated from the tropics to drive ice-sheet melting in the Northern Hemisphere 2,000 to 3,000 years later. The strength of the evidence lies in the recording of the lagged response of ice-sheet melting using the same foraminiferal signal carrier in the same sediment core. The authors’ views are further supported by their observation of the same sequence of events and extent of temperature change during a previous deglaciation about 130,000 years ago.

The sediment cores that Visser *et al.* analysed were also dated using radiocarbon. When compared with evidence from polar ice cores, it seems that warming in the tropics, beginning just over 18,000 years ago, is synchronous with the initial rise in atmospheric CO_2 , as well as with the warming recorded during deglaciation in Antarctica. These findings are somewhat less secure, as they are based on comparisons of data obtained with different dating systems. Nevertheless, they point to deglaciation that was largely driven by tropical events, probably through the reinforcing greenhouse effects of two gases — CO_2 and water vapour. If the Indo-Pacific warm pool warmed during deglaciation, simple thermodynamic and solubility considerations

mean that the ocean would have supplied additional water vapour and CO_2 to the atmosphere, wherein an enhanced greenhouse effect would act as a positive feedback on planetary warming. Antarctic temperatures appear to have responded quickly. Ultimately, this warming led to rapid melting of ice sheets in the Northern Hemisphere several thousand years later.

Inevitably, there are caveats about this work. The foraminiferal Mg/Ca thermometer is a relatively new tool: although the temperature signal is set during the lifespan of the foraminifera, it is not yet clear to what extent it may be altered as their delicate carbonate shells settle through the ocean and are buried at the sea floor. Moreover, there are only a handful of Mg/Ca time series spanning the last deglaciation, and no others from the heart of the warm pool. Palaeoclimate scientists have learned that replication sometimes yields surprising results, and additional cores from the western Pacific should be similarly analysed. Visser *et al.* also suggest that warming of the tropical ocean by 3.5–4 °C could account for much of the 90 p.p.m. rise in atmospheric CO_2 through a direct solubility effect alone (warmer water holds less CO_2). Yet, from other estimates⁷, it seems that during deglaciation the direct effect of temperature on CO_2 levels is smaller, suggesting

that the transfer of carbon between reservoirs was more complex.

Nevertheless, the finding that temperatures in the tropics increased before the onset of melting of the northern ice sheets is consistent with several other studies^{8,9}. And if it is supported by further analyses, it will be highly significant because it suggests a way in which the tropics can amplify a small, direct radiative signal, such as that associated with long-period orbital variations, to produce a large and dynamic climate change. As we head uncertainly into a greenhouse world, we should perhaps expect that some of the climate surprises that await us will emanate from the tropics. ■

Robert B. Dunbar is in the Department of Geological and Environmental Sciences, 325 Braun Hall, Stanford University, Stanford, California 94305-2115, USA.
e-mail: dunbar@stanford.edu

1. Visser, K., Thunell, R. & Stott, L. *Nature* **421**, 152–155 (2003).
2. Petit, J. R. *et al.* *Nature* **399**, 429–436 (1999).
3. Croll, J. *Climate and Time* (Appleton, New York, 1875).
4. Milankovitch, M. in *Handbuch der Klimatologie* (eds Köppen, W. & Geiger, R.) 1–176 (Borntraeger, Berlin, 1930).
5. Lea, D., Mashiotta, T. & Spero, H. *Geochim. Cosmochim. Acta* **63**, 2369–2379 (1999).
6. Hastings, D. W., Russell, A. D. & Emerson, S. R. *Paleoceanography* **13**, 161–169 (1998).
7. Archer, D. *et al.* *Glob. Biogeochem. Cycles* **14**, 1219–1230 (2001).
8. Guilderson, T. P., Fairbanks, R. G. & Rubinstone, J. L. *Mar. Geol.* **172**, 75–89 (2001).
9. Seltzer, G. O. *et al.* *Science* **296**, 1685–1686 (2002).

Molecular systematics

Counting angels with DNA

Mark Blaxter

It is impossible to describe biological diversity with traditional approaches. Molecular methods are the way forward — especially, perhaps, in the form of DNA barcodes.

For 200 years taxonomists have recognized two living species of elephant, the African and the Indian. According to molecular evidence, however, this understanding is wrong^{1,2}. Although the details are still under discussion, two or maybe three groups of ‘African’ elephant are as distinct from each other as either is from the Indian, and so may constitute distinct taxonomic groups — a finding with implications for conservation. This kind of discovery, using molecular markers, epitomizes a sea change occurring in taxonomy. It underlies a proposal, published by Hebert *et al.*³ in *Proceedings of the Royal Society*, for a DNA-based barcoding system for all animal species. This is not a new idea for some taxonomic groups. But Hebert *et al.* now formally suggest that a molecular barcode inventory should be made of known animal taxa, and that this database should become the basis of biodiversity assessment and taxon identification.

There may be over 100 million extant species on Earth, but only a tiny proportion

of them have been described⁴. The scale of the descriptive deficit varies widely for different groups of organisms (Fig. 1). For taxa with large body sizes, such as vertebrates, the catalogue may be essentially complete. For smaller taxa (with bodies 0.5–10 mm long) that is far from the case. For instance, there are some 1 million described species of arthropods (insects and allies), but estimates of actual diversity range from 3 million to 30 million. Only 20,000 species of nematode (roundworms) have been described out of a predicted million or so⁵. For the smallest organisms — the bacteria, archaea and single-celled eukaryotes — we may be ignorant of over 99% of actual diversity⁶. It is not feasible⁷ to catalogue that diversity by traditional methods based on taxon-by-taxon, specimen-by-specimen morphological description. Instead, emerging technologies must be harnessed to the task⁸.

In the 1990s, ideas about the biosphere were revolutionized by surveys of the diversity of bacteria and archaea (the prokaryotes)

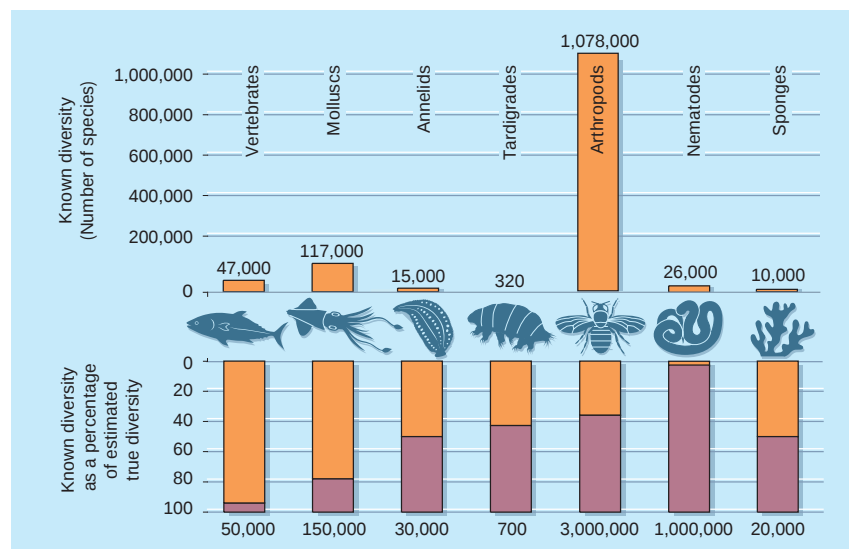


Figure 1 Known and estimated diversity of a selection of animal phyla. The histogram above each representative image indicates known diversity (in terms of number of species). That below represents known diversity as a percentage of estimated true diversity, shown by the number at the bottom. This graphic provides only a keyhole view of biodiversity and the task the systematist faces. It is confined to just a few phyla of the larger eukaryotes (loosely, organisms whose cells contain a nucleus), of which there are many more groups, including the plants and unicellular forms. And then there are the vast numbers of microscopic prokaryotes, consisting of the bacteria and archaea. Molecular barcoding methods may be the only way of charting unknown diversity.

using DNA sequences. All cellular organisms have the highly conserved small subunit ribosomal RNA (SSU) gene, which can be isolated specifically even from bulk samples of 'environmental DNA'. By surveying SSU gene sequences in a sample, it is possible to identify how many different taxa are present, and assess their relationships to previously described and sequenced groups. All environments sampled yield the same pattern: many of the constituent microorganisms come from major groups unknown to traditional microbiology, and the diversity of prokaryotes is probably 100 times higher than was previously expected^{6,9}. Similar SSU gene surveys of communities of eukaryotic microbes, such as marine picoplankton, also suggest that traditional, descriptive methods have sampled only the tip of diversity^{10–12}.

It is now accepted in bacteriology that taxa can be identified using sequence data, and rules of thumb for defining taxa based on sequence difference are developing¹³. The taxa so defined have been termed 'phylo-types'¹⁰ or 'molecular operational taxonomic units'¹⁴. Thus a DNA sequence can be used to both identify and classify an organism, much as a barcode identifies supermarket products.

Hebert *et al.*³ extend this idea to non-microbes, and propose that a database of DNA barcodes for identification of all animal taxa should be established. They test the use of the mitochondrially encoded cytochrome oxidase I (COI) gene in assigning specimens to different taxonomic levels — phylum, order and species. The COI gene

is a good target, as it is present in all animals, in many identical copies per cell, and is known to evolve relatively quickly. It carries sufficient 'signal' to allow differentiation between closely related taxa, and in a survey of moths collected around Guelph, Ontario, the authors show that the COI barcode can in most cases yield a reliable assignment to previously identified and sequenced species. Other insect specimens were correctly assigned to the superfamily level, and Hebert *et al.* claim that in general the approach can identify which phylum a sequence derives from.

This is indeed a promising approach. Many potential barcode sequences already exist in public databases: about 12,000 COI sequences, over 20,000 SSU gene sequences (from all organisms)¹⁵, and more than 50,000 sequences from the 'ribosomal internal transcribed spacer segment' (from higher eukaryotes). Plant systematists have also extensively surveyed the photosynthetic *rbcL* gene¹⁶. My group has used the SSU gene to successfully barcode soil nematode biodiversity¹³, and the large subunit ribosomal RNA gene has been used for identifying freshwater fauna¹⁷. It is not necessary to limit data collection to one gene, as sensitive amplification techniques allow the isolation of several sequences from one specimen. Given technological advance and genome-scale sequencing, sequencing of several barcode genes for large numbers of specimens could be achieved rapidly and cheaply.

A major unresolved issue is how closely these molecular taxa correspond to what



100 YEARS AGO

The unfortunate fatal accident which occurred at the Fulham Public Baths on December 23 serves to show how dangerous an electric shock may be when the conditions are such that really good contact is made. In this case, two bathers were killed by standing up in their baths and putting their hands on a metal rail running along the top of the partition between the baths; on top of this rail ran the iron pipes containing the electric-supply leads. It seems that there was leakage, possibly in a faulty lampholder, to these pipes, which were insufficiently "earthed". The bathers therefore completed the earth through their bodies to the bath itself, and thus received a shock which, in spite of the fact that the pressure could only have been something like 170 volts, had fatal results on account of the very good contacts which existed. The circumstances of the case are altogether exceptional, and there is absolutely no need for users of electric light to take any alarm.

From *Nature* 8 January 1903.

50 YEARS AGO

Gregorio Ricci-Curbastro, inventor of the tensor calculus, was born at Lugo, in Italy, on January 12, 1853. The absolute differential calculus, as he himself called it, gained little attention until Einstein used it for the formulation of general relativity, even though it had reached a mature form by 1895, after some ten years of growth. It was so little thought of, indeed, that in 1901 Ricci was denied the Italian Royal Prize in mathematics on the ground that the calculus was "useful but not essential for the treatment of some mathematical questions". Nevertheless he himself retained a belief in its value... In 1912 Einstein's attention was directed to it by his colleague, Marcel Grossmann, and the outcome was the relativistic theory of gravitation published in 1916... The relativistic principle of covariance, namely, that the general laws of physics can be expressed in a form which is independent of the co-ordinate system, has a meaning only in so far as there exists a way of expressing them in such a form. The Ricci calculus provides a means of doing so... The tensor calculus is fully established as one of the main instruments of modern mathematics, and gives to its inventor a permanent place in the history of the subject.

From *Nature* 10 January 1953.

traditional biologists would recognize as a species. Species concepts are difficult to apply to bacteria and protozoa, where sexual isolation (the cornerstone of the original biological species concept) is not relevant. A bewildering abundance of species concepts have been proposed and, practically, there are many cases that fall between different sets of criteria^{18,19}: molecular taxon concepts may add to the confusion. This problem is properly the domain of systematists, but a useful compromise will be achieved by extensive sampling from within recognized species and across known species 'flocks'; and developing rules of thumb that can be used to associate molecular differences with useful biological categories¹².

Hebert *et al.* approach this problem by showing that anonymous specimens were robustly associated with their correct morphological species, but it is likely that different rules may have to be implemented for different animal groups with different evolutionary rates. Until a comprehensive database is assembled, it is not clear how helpful the system using COI will be: for example, the nematode samples in Hebert and colleagues' data set were not well resolved, and there were significant problems with speciose taxa such as beetles.

Nonetheless, a taxonomic genomics programme should yield huge benefits for biodiversity science and ecology, much as the sequencing of genomes has benefited other areas of biology. As was the case with the initial work of Hebert and colleagues, the first stages of such a programme will need identified specimens to represent known diversity and link barcodes to biology. Subsequently, barcoding of environmental samples or collections of unidentified specimens will reveal novelties: flocks of taxa where one

species was expected, or divergent taxa with possible novel biology. One of the advantages of this approach is that the raw data — the DNA sequence — can be stored in databases and accessed through the Internet for comparison and reanalysis. Another is that debate about what usefully defines a taxon, how many taxa there are and what taxon diversity means will become a data-rich science, rather than resembling theological speculation as to how many angels can dance on the head of a pin. ■

Mark Blaxter is at the Institute of Cell, Animal and Population Biology, University of Edinburgh, Edinburgh EH9 3JT, UK.

email: mark.blaxter@ed.ac.uk

1. Roca, A. L., Georgiadis, N., Pecon-Slaterry, J. & O'Brien, S. J. *Science* **293**, 1473–1477 (2001).
2. Eggert, L. S., Rasner, C. A. & Woodruff, D. S. *Proc. R. Soc. Lond. B* **269**, 1993–2006 (2002).
3. Hebert, P. D. N., Cywinska, A., Ball, S. L. & deWaard, J. R. *Proc. R. Soc. Lond. B* (in the press); www.pubs.royalsoc.ac.uk
4. May, R. M. *Science* **241**, 1441–1449 (1988).
5. Lamshead, J. *Oceanis* **19**, 5–24 (1993).
6. Woese, C. R. *Curr. Biol.* **6**, 1060–1063 (1996).
7. Lawton, J. H. *et al.* *Nature* **391**, 72–75 (1998).
8. Godfray, H. J. C. *Nature* **417**, 17–19 (2002).
9. Pace, N. R. *Science* **276**, 734–740 (1997).
10. Moreira, D. & Lopez-Garcia, P. *Trends Microbiol.* **10**, 31–38 (2002).
11. Lopez-Garcia, P., Rodriguez-Valera, F., Pedros-Alio, C. & Moreira, D. *Nature* **409**, 603–607 (2001).
12. Moon-van der Staay, S. Y., De Wachter, R. & Vaulot, D. *Nature* **409**, 607–610 (2001).
13. Theron, J. & Cloete, T. E. *Crit. Rev. Microbiol.* **26**, 37–57 (2000).
14. Floyd, R., Abebe, E., Papert, A. & Blaxter, M. *Mol. Ecol.* **11**, 839–850 (2002).
15. Wuyts, J., van der Peer, Y., Winkelmans, T. & De Wachter, R. *Nucleic Acids Res.* **30**, 183–185 (2002).
16. Chase, M. W. & Fay, M. F. *Genome Biol.* **2**(4): reviews1012 (2001).
17. Markmann, M. *Entwicklung und Anwendung einer 28S rDNA Sequenzdatenbank zur Ausschlusselung der Artenvielfalt limnischer Meiofauna im Hinblick auf den Einsatz moderner Chiptechnologie*. Thesis, Ludwig Maximilians Univ., Munich (2000).
18. Noor, M. A. F. *Trends Ecol. Evol.* **17**, 153–154 (2002).
19. Adams, B. J. J. *Nematol.* **33**, 153–160 (2001).

example of a fluid so resistant to flow that it 'remembers' its past history and produces forces within it that are proportional to the spatial inhomogeneities in its stretching. This resistance tends to make the incursions into these materials narrow and crack-like, similar to the hairline fractures in a brittle material. Experimentally, this kind of behaviour has been seen in wet clays² and in dilute solutions containing long-chain polymers³.

Levermann and Procaccia's technique builds on the work of Hastings and Levitov⁴, who developed methods based on the theory of analytic functions of complex variables to describe situations in which the retreating fluid was an ordinary, 'newtonian' viscous fluid. The pressure in this kind of fluid obeys the partial differential equation called Laplace's equation. Mathematicians would say that the pressure is a harmonic function. Laplace's equation provides a precise statement of the condition that the pressure within a region varies as smoothly as it can, given its values at the boundaries of the region. Complex-variable methods provide a natural framework and convenient language to describe any situation possessing planar geometry, and particularly harmonic functions. From the coordinates x and y in a plane, the complex variable $z = x + iy$ (where i is the square root of -1) can be defined. As if by magic, the maths tells us that any analytical function of z obeys Laplace's equation. These functions can then be used to build a computer model to calculate the pressure in a system.

For an elastic, rather than a viscous, material, the equation to be solved is more difficult. Levermann and Procaccia have used complex-variable analysis to solve the equations for the elastic material. They obtain the displacement of the material and the forces within it, given only the shape of the crack.

But how can the shape of the crack be determined? Here is the novelty of Levermann and Procaccia's work. In the past, it has mostly been assumed that the motion at the boundary between the two materials is smooth. Instead, Levermann and Procaccia divide the motion into little bursts of activity, with each burst localized in a small region of the interface between the two materials. The probability that a burst will happen depends on the forces in that region of the material boundary. Bursting could be explained by the material being so dirty that the motion takes the form of successive events of sticking and slipping, a chaotic jittering motion. This same motion was assumed by Witten and Sander⁵. Their idea was later applied to viscous flow, and then extended⁶ to problems of fracture in amorphous solids.

To make their calculation predictive, Levermann and Procaccia choose a particular model for calculating the probability of a burst at each point on the interface. They roll the computer's dice to figure out where the

Materials science

Bursting apart

Leo P. Kadanoff

When a low-viscosity fluid is injected into an elastic material, it forces its way through by making slender cracks, in a random, fractal pattern. The spreading of the cracks can be modelled through a series of 'bursts'.

Nature abounds in branched objects possessing random, fractal structures: the pattern of air vessels in our lungs, or the branched structure of a river and of the streams and rivulets that feed it. Such patterns arise when the branching process occurs easily and is sensitive to small, random features in the environment. A laboratory example is the pressure-driven expansion of one low-viscosity fluid into another, more viscous fluid. As the fingers of low-viscosity material push outwards, they create high pressure-gradients, and their motion becomes unstable. As a result, little bits of

'dirt', or some other randomness, can considerably affect the flow. The dirt effects show up both in the details of the pattern and also in its gross appearance.

Writing in *Physical Review Letters*, Anders Levermann and Itamar Procaccia¹ describe a calculational procedure that can predict the shape of these patterns. The authors have extended the theories and simulations developed for two-dimensional viscous fluids to the extreme case in which the retreating medium flows elastically. Elastic materials such as rubber resist stretching and deformation. They may be considered to be the extreme

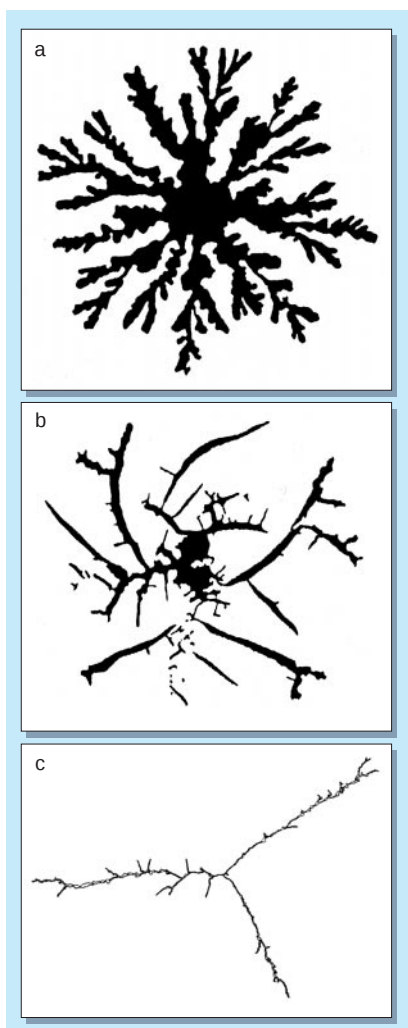


Figure 1 Cracking up. For a fluid injected into viscous low-density mud (a), the measured pattern² of its spread looks very different from the pattern in higher-density, elastic clay (b): in the elastic medium, the pattern is much less full and less branched. Levermann and Procaccia¹ have calculated a theoretical profile (c) for a low-viscosity liquid injected into an elastic medium. Their simulation is closer to the measured elastic pattern than the viscous one, but shows still less branching. (Graphics derived from refs 1, 2.)

next burst will occur, move the interface with that burst, recalculate the stresses and strains and then calculate the probability of burst activity in this new configuration. A complete simulation is shown in Fig. 1, and, although not exactly right, the pattern is indeed reminiscent of those seen in experiments² with elastic materials. For now, the pictures are all we have; more quantitative tests of the model should be developed.

As Levermann and Procaccia point out with commendable candour, the applicability of their method to any particular example of fluid flow depends on several assumptions that are not obviously correct and which should be tested in future work. For example,

the retreating material must be truly elastic for Levermann and Procaccia's method to work, whereas almost all materials in fact show a mixture of elastic and viscous behaviour. Moreover, for any given fluid it is not clear that the particular algorithm used to calculate the bursts is the right one — other workers have disagreed over which equations describe the motion of the interface. The bursting nature of the model might, in any case, not be a correct representation of the situation. Nevertheless, this work represents a promising starting point, and

deserves to be checked and extended. ■

Leo P. Kadanoff is at The James Franck Institute, University of Chicago, 5640 South Ellis Avenue, Chicago, Illinois 60637, USA.

e-mail: leop@uchicago.edu

1. Levermann, A. & Procaccia, I. *Phys. Rev. Lett.* **89**, 234501 (2002).
2. Lemaire, E., Levitz, P., Daccord, G. & van Damme, H. *Phys. Rev. Lett.* **67**, 2009–2012 (1991).
3. Zhao, H. & Maher, J. V. *Phys. Rev. E* **47**, 4278–4283 (1993).
4. Hastings, M. B. & Levitov, L. S. *Physica D* **116**, 244–252 (1998).
5. Witten, T. A. Jr & Sander, L. M. *Phys. Rev. Lett.* **47**, 1400–1403 (1981).
6. Barra, F., Hentschel, H. G. E., Levermann, A. & Procaccia, I. *Phys. Rev. E* **65**, 045101 (2002).

Physiology

Cost-free longevity in mice?

Gordon J. Lithgow and Matthew S. Gill

Studies of worms have revealed hundreds of proteins that, when mutated, extend lifespan. Can this work tell us anything about mammalian ageing? A look at the effects of one such protein on lab mice suggests that it can.

Just over five years ago, molecular geneticists working with the microscopic worm *Caenorhabditis elegans* made an intriguing discovery¹ about a protein that, when its activity is reduced, doubles the animal's lifespan². They found that the protein — DAF-2 — is structurally similar to the human receptor proteins that allow cells to respond to insulin or insulin-like growth factors (IGF). Moreover, a similar receptor influences the lifespan of the worm's distant cousin, the fruitfly *Drosophila melanogaster*³. So could the insulin or IGF receptors also influence human longevity? If so, this would have significant implications for the study and treatment of age-related diseases. On page 182 of this issue, Holzenberger and colleagues report that reducing the levels of the receptor for a subtype of IGF (IGF-I) significantly lengthens the lifespan of laboratory mice⁴. This exciting finding hints that very different sorts of animals, from worms to mammals, age in similar ways. Moreover, it seems that IGF and its receptor are prime candidates for determinants of human longevity.

Ageing is the single largest problem facing biomedicine in developed countries. So it is not surprising that any discovery in the field leads to endless cogitation — none more so than the identification over the past decade of hundreds of *C. elegans* genes that, when mutated, greatly increase lifespan. Each new finding has prompted the same question: could studies of these worm genes reveal general features of ageing that might be applied in combating human age-related diseases?

Hopes have been high ever since the nature of the DAF-2 protein was identified. In *C. elegans*, DAF-2 detects the presence of both insulin-like and IGF-like peptides at the surface of cells, and causes a wave of

intracellular signalling activity that can affect development, fertility, metabolism and the response to stress, as well as lifespan. In mammals, however, the situation is complicated by the existence of distinct insulin and IGF receptors, with the insulin receptor being generally associated with metabolic regulation, and the IGF-I receptor being responsible for growth (Fig. 1, overleaf). Given this additional complexity, it has been important to find out which signalling pathway — insulin-activated or IGF-activated — is involved in regulating mammalian lifespan.

Clues have been emerging from other genetic and environmental interventions that extend the lifespan of mice⁵ (Fig. 1). For instance, some dwarf mouse strains that show abnormal development of the pituitary gland have low levels of growth hormone (a positive regulator of IGF-I), and are long-lived. Similarly, inactivating the growth-hormone receptor lowers IGF-I levels, causing dwarfism, and also extends mouse lifespan. Furthermore, restricting the calorie intake of mice from birth produces low IGF-I levels, dwarf animals, and an increase in lifespan. In all of these cases, longevity is associated with a reduction in IGF-I levels (and hence decreased signalling through the IGF-I receptor) throughout life, with a corresponding deficit in body size. This contrasts, however, with reports that restoring normal IGF-I levels in elderly people, by means of growth-hormone-replacement therapy, may result in significant health benefits⁶.

Genetically altered mice have proved to be essential tools for unravelling the complex and interacting roles of insulin and IGF-I⁷. Using this approach, Holzenberger *et al.*⁴ now confirm the link between IGF-I and ageing in mammals, and show that it is

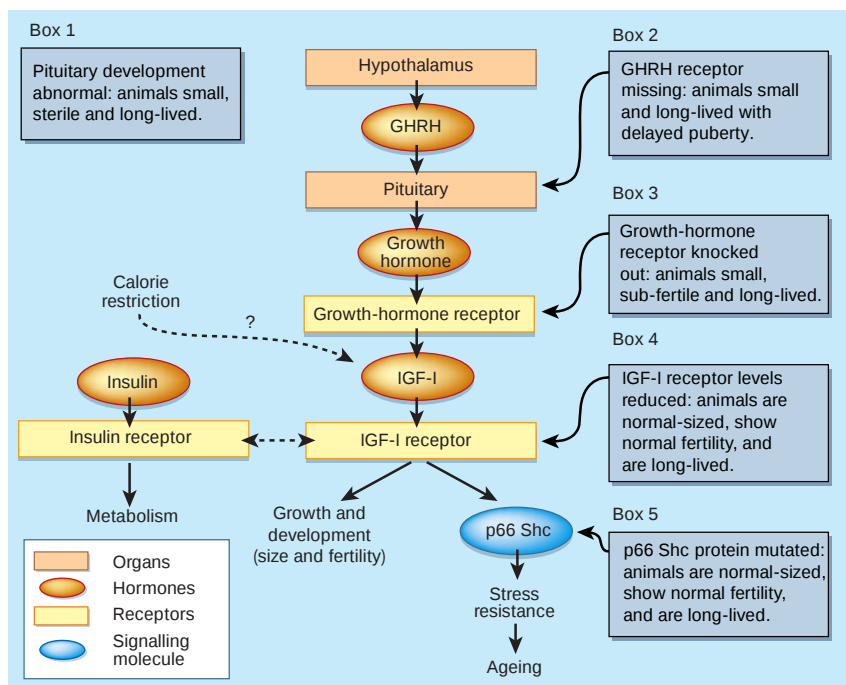


Figure 1 Genetic dissection of the hormonal determinants of mammalian lifespan. Many lab mice that live longer than normal have genetic defects that affect the production (boxes 1 and 2) or action (box 3) of growth hormone, thereby influencing both body size and fertility, as can be seen from this signalling pathway. Holzenberger *et al.*⁴ have shown that reducing the levels of the receptor for insulin-like growth factor-I (IGF-I) extends lifespan and enhances resistance to oxidative stress in mice (box 4), without any apparent detrimental effect on fertility and body size. Similarly, mice with a mutation in the signalling molecule p66 Shc are long-lived and resistant to oxidative stress⁸ (box 5). Holzenberger *et al.* also show that p66 Shc is altered in IGF-I-receptor-deficient mice, indicating that this signalling pathway is important for lifespan extension. Calorie restriction may also increase lifespan by affecting IGF-I signalling. GHRH, growth-hormone-releasing hormone.

possible to separate the effects of IGF-I signalling on lifespan and body size. They find that if both copies of the gene encoding the IGF-I receptor — *igf1r* — are knocked out, mice die at birth with an array of developmental abnormalities.

However, knocking out only one copy of the gene (producing *igf1r*^{+/-}, or 'heterozygous', mice) leads to an extended lifespan. This is one of the more difficult life-history traits to measure, as it is sensitive to minor environmental stresses and strain-specific differences, and shows considerable variability even among genetically identical animals. Nonetheless, Holzenberger *et al.* were able to show that the *igf1r*^{+/-} mice live, on average, 26% longer than their wild-type counterparts. This seems to reflect a decrease in the rate of ageing rather than in the frequency of a specific disease, as the heterozygous and wild-type animals died of the same spectrum of diseases. The authors hint that another loss-of-function mutation in the IGF-I receptor also extends lifespan, and, if true, this will go a long way to supporting the main finding.

Significantly, the *igf1r*^{+/-} mice also grow and develop more or less normally, reaching puberty on schedule and producing the same number of pups, at the same rate, as wild-

type mice. They have normal metabolic rates, body temperatures, feeding behaviour, activity and fertility, and, critically, their body size is also normal⁴. This will help in identifying the crucial factors that confer longevity (Fig. 1). It seems that a mouse does not have to be sub-fertile, metabolically compromised or small to age slowly.

Given these findings⁴, can we now exclude a role for insulin in mammalian ageing? Not quite. The insulin and IGF-I signalling pathways communicate extensively: for instance, although IGF-I acts exclusively through the IGF-I receptor, there is evidence that insulin can bind to both this receptor and the insulin receptor. Moreover, hybrid insulin-IGF-I receptors occur in some tissues⁷. So the story of the long-lived *igf1r*^{+/-} mouse will not end here. To understand how IGF-I affects lifespan, it will be important to examine the effects on specific tissues and to identify downstream signalling events.

What does all this tell us about the causes of mammalian ageing? Reduced signalling from insulin-like peptides in *C. elegans* probably extends lifespan by enhancing resistance to intrinsic stresses, such as oxygen radicals. The same might be true of mammals. The idea that ageing results from the accumulation of oxygen-radical-induced molecular

and cellular damage has been around since the 1950s. So it is significant that *igf1r*^{+/-} mice and *igf1r*^{+/-} cells in culture are resistant to attack by oxygen radicals⁴.

The normal IGF receptor might decrease resistance to oxidative stress in several ways. It is thought, for instance, to work through two intracellular signalling pathways, one of which involves two forms of the Shc protein (p66 and p52), and the other the enzyme phosphatidylinositol-3-OH kinase. p66 Shc has already been studied in an ageing context — mutating this protein extends mouse lifespan by 30% and enhances resistance to oxidative stress⁸. In *igf1r*^{+/-} mice, both forms of Shc as well as phosphatidylinositol-3-OH kinase are affected, so both pathways should be studied further.

An interesting subplot to Holzenberger and colleagues' publication is the debate over the evolutionary origins of ageing. Most researchers agree that organismal ageing does not follow a predetermined series of steps in the same way as does, say, programmed cell death. Ageing is thought not to have any direct effect on evolutionary fitness, but to be a by-product of the intense pressure of natural selection on growth and early reproduction. The result is that longevity usually comes at a cost. But where is this cost in the *igf1r*^{+/-} mice? There is evidence of a defect in B-cell development in males, but the measured life-history traits of the females (which tended to live still longer) seem normal⁴. Long-lived mutant worms also appear to pull off this trick of cost-free longevity, if only in specific laboratory environments. The question of whether *igf1r*^{+/-} mice have compromised evolutionary fitness will be of interest to those who ponder not only the mechanisms but also the origins of ageing.

This all suggests a bright future for ageing research. It appears that studies of simple animals are portentous, and an increasing number of longevity mutations found in invertebrates are now waiting to be tested in rodents. One wonders how long it will be before human ageing can be counted among the many biological effects of IGF-I.

Gordon J. Lithgow and Matthew S. Gill are at The Buck Institute, 8001 Redwood Boulevard, Novato, California 94949, USA.
e-mails: glithgow@buckinstitute.org
mgill@buckinstitute.org

- Kimura, K. D., Tissenbaum, H. A., Liu, Y. & Ruvkun, G. *Science* **277**, 942–946 (1997).
- Kenyon, C., Chang, J., Gensch, E., Rudner, A. & Tabtiang, R. *Nature* **366**, 461–464 (1993).
- Tatar, M. *et al. Science* **292**, 107–110 (2001).
- Holzenberger, M. *et al. Nature* **421**, 182–187 (2003); advance online publication, 4 December 2002 (doi:10.1038/nature01298).
- Bartke, A. *et al. J. Gerontol. A Biol. Sci. Med. Sci.* **56**, B340–B349 (2001).
- Carter, C. S., Ramsey, M. M. & Sonntag, W. E. *Trends Genet.* **18**, 295–301 (2002).
- Nakae, J., Kido, Y. & Accili, D. *Endocrinol. Rev.* **22**, 818–835 (2001).
- Migliaccio, E. *et al. Nature* **402**, 309–313 (1999).

Rhodopsin dimers in native disc membranes

Neat rows of paired photon receptors are caught on camera in their natural state.

In vertebrate retinal photoreceptors, the rod outer-segment disc membranes contain densely packed rhodopsin molecules for optimal light absorption and subsequent amplification by the visual signalling cascade¹, but how these photon receptors are organized with respect to each other is not known. Here we use infrared-laser atomic-force microscopy to reveal the native arrangement of rhodopsin, which forms paracrystalline arrays of dimers in mouse disc membranes. The visualization of these closely packed rhodopsin dimers in native membranes gives experimental support to earlier inferences about their supramolecular structure^{2,3} and provides insight into how light signalling is controlled.

When activated by a single photon, rhodopsin induces dissociation of the subunits of transducin molecules, which are heterotrimeric G proteins that amplify the light signal. The structure of rhodopsin, which has been solved⁴, indicates that it is a prototypical member of subfamily A of G-protein-coupled receptors (GPCRs), which represents about 90% of all GPCRs. The oligomerization² of these GPCRs may be important for their own regulation and for interaction with their cognate G proteins.

The probability of single-photon absorption by rhodopsin molecules is increased by their packing arrangement on the tightly stacked rod outer-segment disc membranes⁵. To investigate this organization of rhodopsin, we isolated rod outer-segment disc membranes from mouse retinas, using a protocol that preserves the full biological activity and organization of rhodopsin in the lipid membrane⁶ (Y.L. *et al.*, manuscript submitted).

Atomic-force microscopy revealed single-layered disc membranes that were 7–8 nm thick and circular in shape (diameter, 0.9–1.5 μm). Occasionally, complete discs with two stacked membranes were also seen; the thickness of these was 16–17 nm. The most frequently observed surface type (Fig. 1, region 1) was 7.8 ± 0.2 nm thick ($n=55$) and had a markedly textured topography consisting of densely packed lines. Lipid bilayers had an unstructured topography (Fig. 1, region 2), a height of 3.7 ± 0.2 nm ($n=86$), and were often seen at the borders of disc membranes. The mica support (Fig. 1, region 3) had no structural features, as expected.

The highly textured surface was assigned as cytoplasmic in view of its relative stiff-

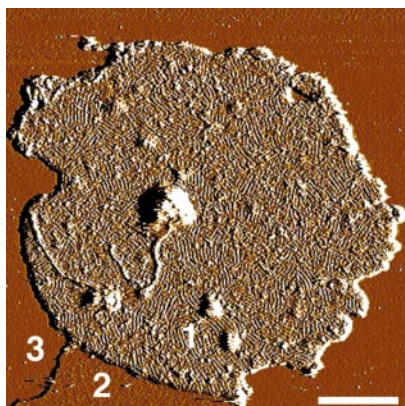


Figure 1 Deflection image of a native eye-disc membrane adsorbed on mica, visualized by atomic-force microscopy (Nanoscope Multimode, Digital Instruments). Three different surface types are evident: 1, the cytoplasmic side of the disc membrane; 2, lipid; and 3, mica. To avoid the formation of opsin, the chromophore-depleted form of rhodopsin⁶, membrane samples were never exposed to light. After adsorption of osmotically shocked disc membranes onto mica, their topography was measured in buffer solution (20 mM Tris-HCl (pH 7.8), 150 mM KCl and 25 mM MgCl_2). Scale bar, 200 nm.

ness. At high magnification, the topography (Fig. 2a, c) revealed rows of rhodopsin pairs densely packed in paracrystalline arrays. Packing densities were 30,000–55,000 rhodopsin monomers per μm^2 , with an average density of $48,300 \pm 8,300$ monomers per μm^2 .

We deduced the dimensions of the rhodopsin pairs from the angularly averaged

powder-diffraction pattern (Fig. 2a, inset, and b, arrows): the innermost arc peaks at $(8.4 \text{ nm})^{-1}$, which results from regularly packed double rows of protrusions; the next ring, at $(4.2 \text{ nm})^{-1}$, reflects the second-order of the double-row repeat, and the axial repeat of the paired rhodopsins that form these rows yields a third ring at $(3.8 \text{ nm})^{-1}$. From real space measurements, we found the distance between the protrusions within each pair to be 3.8 ± 0.2 nm ($n=40$) and the angle between the lattice vectors to be $85 \pm 2^\circ$ ($n=8$).

From these dimensions, the highest possible packing density is 62,900 rhodopsin monomers per μm^2 . At higher magnification (Fig. 2c), it can be seen that almost all rhodopsin molecules are present in rows of dimers, with a few monomers and some single rhodopsin pairs that have broken away from the rows. This direct demonstration of distinct, densely packed rows of dimers at high resolution is consistent with the proposed dimeric form of native GPCRs that has been inferred from biochemical and pharmacological analysis² and from evolutionary trace analysis³.

The oligomeric organization of GPCR signalling molecules has important implications for G-protein recognition, binding kinetics, signal amplification and signal termination. The shape and dimensions of the paracrystalline unit cell set stringent boundaries for the dimer configuration, the

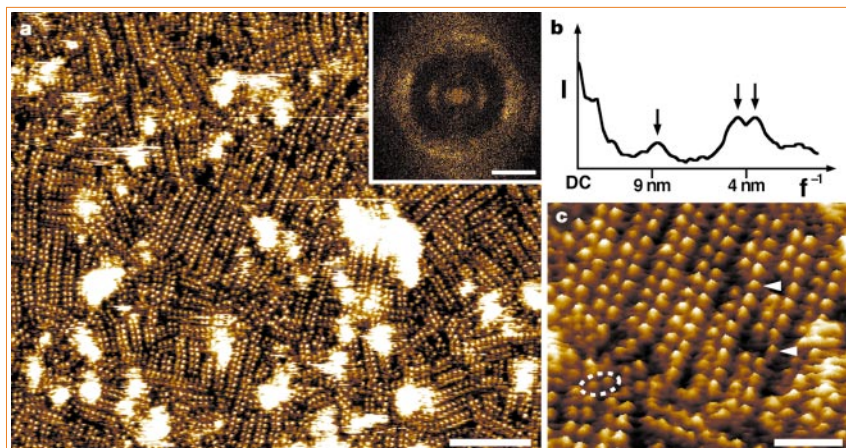


Figure 2 Organization and topography of the cytoplasmic surface of rhodopsin. **a**, Topograph obtained using atomic-force microscopy, showing the paracrystalline arrangement of rhodopsin dimers in the native disc membrane. Inset, arcs in the calculated powder-diffraction pattern reflect the regular arrangement of rhodopsin in the membrane. **b**, Angularly averaged powder-diffraction pattern, showing peaks at $(8.4 \text{ nm})^{-1}$, $(4.2 \text{ nm})^{-1}$ and $(3.8 \text{ nm})^{-1}$. **c**, Magnification of a region of the topograph in **a**, showing rows of rhodopsin dimers, as well as individual dimers (inside dashed ellipse), presumably broken away from one of the rows, and occasional rhodopsin monomers (arrowheads). The rhodopsin molecules protrude from the lipid bilayer by 1.4 ± 0.2 nm ($n=111$). The topograph in **c** is shown in relief, tilted by 5° . Vertical brightness ranges: 1.6 nm (**a** and **c**). Scale bars: **a**, 50 nm; inset, $(5 \text{ nm})^{-1}$; **c**, 15 nm.

intra-dimer contacts and the formation of oligomers that consist of greater numbers of units (Protein Data Bank accession number, 1N3M).

Dimitrios Fotiadis*, **Yan Liang†**,
Slawomir Filipek‡, **David A. Saperstein†**,
Andreas Engel*, **Krzysztof Palczewski†§||**

*M. E. Müller Institute for Microscopy,
Biozentrum, University of Basel, Basel 4056,
Switzerland

Departments of †Ophthalmology, §Pharmacology
and ||Chemistry, University of Washington, Seattle,
Washington 98195, USA

e-mail: palczewski@u.washington.edu

‡International Institute of Molecular and Cell
Biology, and Department of Chemistry, University
of Warsaw, Warsaw 02-109, Poland

1. Papermaster, D. S. *Invest. Ophthalmol. Vis. Sci.* **43**, 1300–1309 (2002).
2. Rios, C. D., Jordan, B. A., Gomes, I. & Devi, L. A. *Pharmacol. Ther.* **92**, 71–87 (2001).
3. Dean, M. K. et al. *J. Med. Chem.* **44**, 4595–4614 (2001).
4. Palczewski, K. et al. *Science* **289**, 739–745 (2000).
5. Molday, R. S. *Invest. Ophthalmol. Vis. Sci.* **39**, 2491–2513 (1998).
6. McBee, J. K., Palczewski, K., Baehr, W. & Pepperberg, D. R. *Progr. Retin. Eye Res.* **20**, 469–529 (2001).

Competing financial interests: declared none.

Coulomb fission

Rayleigh jets from levitated microdroplets

Electrified droplets are generated in thunderstorm clouds, as well as in technological applications such as ink-jet printing and electrospray ionization, but they become unstable when charged beyond the Rayleigh limit¹. Here we record the dynamics of the disintegration process by examining levitated droplets under high-speed microscopy. These images may help to explain one of the oldest unsolved problems in experimental and theoretical physics.

Lord Rayleigh showed that the spherical shape of a drop of radius a_0 , surface tension σ and charge Q , remains stable as long as the fissility $X = Q^2 / (64\pi^2 \epsilon_0 \sigma a_0^3)$ does not exceed unity¹. As X approaches unity, the quadrupole deformation is the first to become unstable; when X increases beyond unity, however, an instability occurs that is associated with the formation of fine jets

(known as Rayleigh jets) “whose fineness, however, has a limit”¹. Although this conjecture has since been examined both experimentally^{2–4} and theoretically^{5,6}, the mechanics of the break-up of these charged droplets and the details of the jets’ fineness remain unclear.

We used high-speed microscopic images to observe the disintegration of droplets charged to the Rayleigh limit and the production of Rayleigh jets. In our experimental set-up, a charged droplet of ethylene glycol is produced by using a piezo-driven nozzle and is suspended in an electrodynamic levitator; the droplet is illuminated by an unfocused He–Ne laser and the mass/charge ratio and the radius of the droplet are determined in real time by analysis of the scattered light^{7,8}.

At the moment of injection, the droplet radius, a_0 , is 58 μm and its charge is about 3.3 picocoulombs. It subsequently shrinks through evaporation of neutral molecules, reaching the Rayleigh limit of stability at a radius of about 24 μm . Just before this point, the amplitude of the quadrupole-

shaped oscillation of the droplet increases drastically⁹, triggering a signal to a fast flashlamp that fires after a predefined delay, Δt , and is recorded by a digital microscope as a still image of the droplet. This process is repeated with subsequent droplets, using increasing delay times.

These microscopic images of the disintegration process of our levitated microdroplets are reproduced in Fig. 1a–f. During the first 150 μs , the droplet stretches from a sphere into an ellipsoid, as predicted by Rayleigh (Fig. 1a). It then develops two sharp tips on the poles of the ellipsoid (Fig. 1b). Almost instantaneously after tip formation, a fine jet of liquid is ejected from both tips in opposite directions (Fig. 1c). This jet later disintegrates into fine droplets (Fig. 1d) that are repelled from the mother droplet by Coulomb repulsion. The tip edges disappear after the ejection of the jet (Fig. 1e), and the barrel-shaped parent droplet then contracts until it regains spherical symmetry after about 210 μs (Fig. 1f).

During the jet’s disintegration, roughly 100 small daughter droplets are formed, which carry 33% of the total charge and constitute about 0.3% of the mass of the mother droplet. The diameter of the jet is determined from higher-resolution images to be 1.5 μm : the droplets generated during disintegration are of roughly the same size and are themselves close to a fissility of $X_{\text{frag}} = 1$ and so presumably undergo a Rayleigh instability soon afterwards.

As this behaviour is independent of the levitator voltage, we argue that it must correspond closely to that of free, highly charged droplets. But, in contrast to Lord Rayleigh’s prediction, we observed the jets at a fissility of unity, indicating that renewed investigation will be necessary to explain the complex hydrodynamics of this century-old problem.

Denis Duft, Tobias Aechtzehn, Rene Müller, Bernd A. Huber*, Thomas Leisner

Institut für Physik, Technische Universität Ilmenau,
Postfach 100565, 98684 Ilmenau, Germany

e-mail: thomas.leisner@TU-Ilmenau.de

*Permanent address: Centre Interdisciplinaire de
Recherche Ions Lasers, rue Claude Bloch,
14070 Caen Cedex 5, France

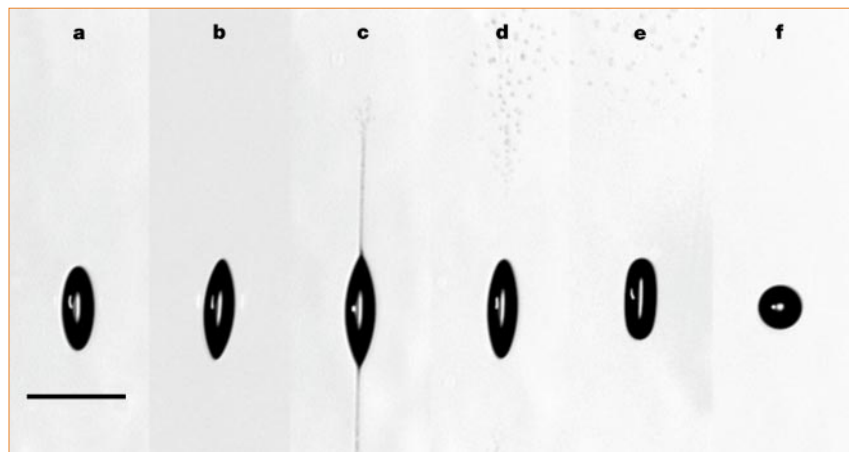


Figure 1 High-speed imaging of the disintegration of a levitated droplet charged to the Rayleigh limit. The droplet (radius, 24 μm) is imaged on a vertical charge-coupled-device array to determine and control its vertical position in the levitator, and on a photomultiplier that detects instability-onset, quadrupole-shaped oscillations of the droplet. These oscillations trigger a signal to a flashlamp that fires after a predefined delay, Δt . The image is observed through a microscope with a long working distance (Mitutoyo). **a–f**, Microscopic images taken at Δt values (in μs) of: **a**, 140; **b**, 150; **c**, 155; **d**, 160; **e**, 180, and **f**, 210. The droplet changes from a sphere to an ellipsoid (**a**), tips appear at the poles (**b**) and a fine jet of liquid is ejected from each tip (**c**); the jets disintegrate (**d**) and the elliptical droplet reassumes a spherical shape (**e**, **f**). Further experimental details are available from the authors. Scale bar, 100 μm .

1. Rayleigh, Lord Phil. Mag. **14**, 184–186 (1882).
2. Taylor, G. I. Proc. R. Soc. Lond. A **280**, 383–397 (1964).
3. Smith, J. N., Flagan, R. C. & Beauchamp, J. L. J. Phys. Chem. A **106**, 9957–9967 (2002).
4. Davis, E. J. & Schweiger, G. *The Airborne Microparticle* (Springer, Heidelberg, 2002).
5. Tsampoulou, J. A., Akyas, T. R. & Brown, R. A. Proc. R. Soc. Lond. A **401**, 67–88 (1985).
6. Pashkevich, V. V., Krappe, H. J. & Wehner, J. Z. Phys. D **40**, 338–340 (1997).
7. Krämer, B. et al. J. Chem. Phys. **111**, 6521–6527 (1999).
8. Schwell, M. et al. J. Phys. Chem. A **104**, 6726–6732 (2000).
9. Duft, D., Leibus, H., Huber, B. A., Guet, C. & Leisner, T. Phys. Rev. Lett. **89**, 0845031–0845034 (2002).

Competing financial interests: declared none.

Econophysics

Master curve for price-impact function

The price reaction to a single transaction depends on transaction volume, the identity of the stock, and possibly many other factors. Here we show that, by taking into account the differences in liquidity for stocks of different size classes of market capitalization, we can rescale both the average price shift and the transaction volume to obtain a uniform price-impact curve for all size classes of firm for four different years (1995–98). This single-curve collapse of the price-impact function suggests that fluctuations from the supply-and-demand equilibrium for many financial assets, differing in economic sectors of activity and market capitalization, are governed by the same statistical rule.

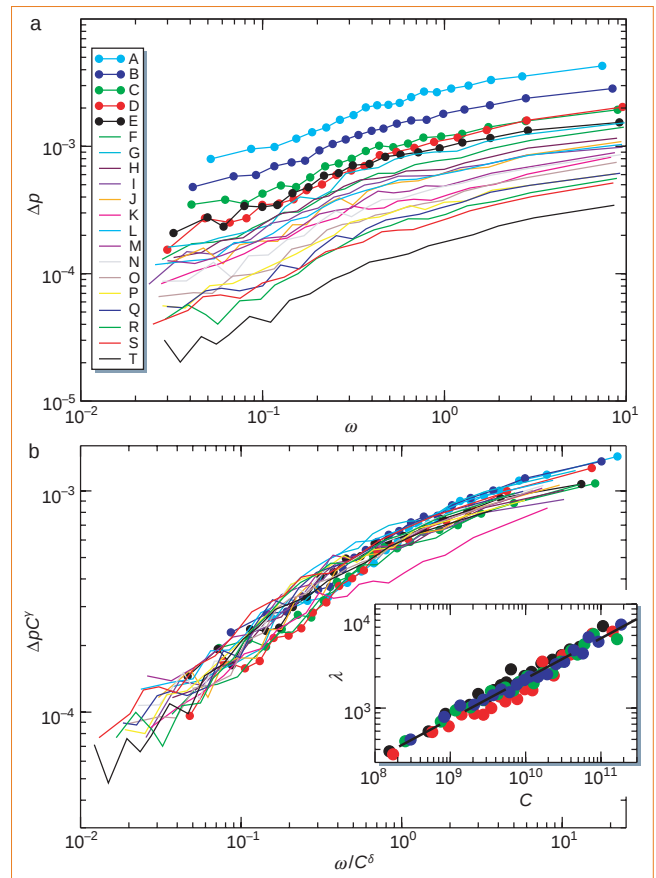
Our results complement previous efforts^{1–9} by using huge amounts of data, by looking at the short-term response to a single trade, and by measuring time in units of transactions rather than in seconds. We used the Trade and Quote database as our data source and studied the 1,000 largest firms on the New York Stock Exchange in 1995–98, by analysing roughly 113 million transactions and 173 million quotes, and investigating the shift in the mid-quote price caused by the most recent transaction.

For each transaction of volume ω made at time t , we observe two cases. First, when the next event is a quote revision, we compare the next quote to the previous quote and compute the difference in the logarithm of the mid-quote price; if the logarithm of the mid-quote price is $p(t)$, we compute the price shift as $\Delta p(t_{i+1}) = p(t_{i+1}) - p(t_i)$, where t_i is the time of the previous quote and t_{i+1} is the time of the immediately following quote.

Second, when the next event is a new transaction, we set the price shift, $\Delta p(t_{i+1})$, to zero. We then investigate the average price shift as a function of the transaction size, ω , doing this separately for buying and selling. The transactions are classified as being initiated by a buyer or seller by using the Lee and Ready algorithm¹⁰. In order to put all stocks on roughly the same footing, we normalize the transaction size by dividing by the average value for each stock in each year.

To understand how market capitalization affects price impact, we group the 1,000 stocks of our sample into 20 groups. The groups are ordered by market capitalization, and the number of stocks in each group is selected to maintain roughly the same number of transactions in each group. We then bin each transaction on the basis of size, and plot the average price impact against the transaction size for each group.

Figure 1 Scaling of the price-impact curves of 1,000 stocks traded on the New York Stock Exchange. **a**, Price shift, Δp , plotted against normalized transaction size, ω , for buyer-initiated trades for 20 groups of stocks, A–T, sorted by market capitalization in 1995. The mean market capitalization increases from group A to group T. **b**, Price shift plotted against transaction size for buy orders in 1995, renormalized as described in the text to make the data collapse roughly onto a single curve. Inset, the liquidity, λ , is shown as a function of mean capitalization, C , of each group of stocks for 1995 (black), 1996 (green), 1997 (blue) and 1998 (red). The black dashed line is the power-law best fit for all points.



The results obtained for 1995 use all of the available data (Fig. 1a). On a double-logarithmic scale, the slope of each curve varies from roughly 0.5 for small transactions in higher-capitalization stocks, to about 0.2 for larger transactions in lower-capitalization stocks. When we repeat this for the years 1996–98, the results are similar, except that the slopes become slightly flatter with time, ranging from roughly 0.4 to 0.1 in 1998.

Higher-capitalization stocks tend to have smaller price responses for the same normalized transaction size. To explain this observation, we carried out a best fit of the impact curves for small values of the normalized transaction size with the functional form $\Delta p = \text{sign}(\omega) |\omega|^\beta / \lambda$, where λ is the liquidity and $\text{sign}(\omega)$ is $+1$ or -1 for buying and selling, respectively. For all four years, the liquidity of each group increases as roughly $C^{0.4}$, where C is the average market capitalization of each group (Fig. 1b, inset).

We make use of this apparent scaling to collapse the data shown in Fig. 1a into a single curve. We assume that $\Delta p(\omega, C) = C^{-\gamma} f(\omega C^\delta)$, where γ and δ are constants, and we rescale the ω and Δp axes of each group according to the transformations $\omega \rightarrow \omega / C^\delta$ and $\Delta p \rightarrow \Delta p C^\gamma$. We then search for the values of δ and γ that place all of the points most accurately on a single curve. In all the years that we investigated,

there is a clear minimum for $\delta \approx \gamma \approx 0.3$. The resulting rescaled price-impact curves for buys in 1995 are shown in Fig. 1b.

We have investigated demand-and-supply fluctuations in a way that is complementary to the traditional approach in economics. The mechanism for making transactions has recently been theoretically modelled by assuming that order placement and cancellation are largely random¹¹, which results in predictions of price impact that are qualitatively consistent with those made here. Our findings show that it is important to model financial institutions in detail, and that it is may be more useful to model human behaviour as random rather than rational for some purposes.

Fabrizio Lillo*, **J. Dooyne Farmer†**,
Rosario N. Mantegna*‡

**Istituto Nazionale per la Fisica della Materia, and
‡Dipartimento di Fisica e Tecnologie Relative,
Università di Palermo, 90128 Palermo, Italy*

e-mail: mantegna@unipa.it

*†Santa Fe Institute, Santa Fe,
New Mexico 87501, USA*

- Hasbrouck, J. *Handbook Stat.* **14**, 647–692 (1996).
- Hausman, J. A. & Lo, A. W. *J. Finan. Econ.* **31**, 319–379 (1992).
- Chan, L. K. C. & Lakonishok, J. *J. Finance* **50**, 1147–1174 (1995).
- Dufour, A. & Engle, R. F. *J. Finance* **55**, 2467–2498 (2000).
- Farmer, J. D. *Slippage 1996* (Predictions Co. Tech. Rep., Santa Fe, New Mexico, 1996); <http://www.predict.com/jdt/slippage.pdf>
- Torre, N. *BARRA Market Impact Model Handbook* (BARRA,

- Berkeley, California, 1997).
7. Kempf, A. & Korn, O. *J. Finan. Mark.* **2**, 29–48 (1999).
 8. Chordia, T., Roll, R. & Subrahmanyam, A. *J. Finan. Econ.* **65**, 111–130 (2002).
 9. Plerou, V., Gopikrishnan, P., Gabaix, X. & Stanley, H. E. *Phys. Rev. E* **66**, 027104 (2002).
 10. Lee, C. M. C. & Ready, M. J. *J. Finance* **46**, 733–746 (1991).
 11. Daniels, M., Farmer, J. D., Iori, G. & Smith, D. E. Preprint <http://xxx.lanl.gov/cond-mat/0112422> (2001).

Competing financial interests: declared none.

Econophysics

Two-phase behaviour of financial markets

Buying and selling in financial markets is driven by demand, which can be quantified by the imbalance in the number of shares transacted by buyers and sellers over a given time interval. Here we analyse the probability distribution of demand, conditioned on its local noise intensity Σ , and discover the surprising existence of a critical threshold, Σ_c . For $\Sigma < \Sigma_c$, the most

probable value of demand is roughly zero; we interpret this as an equilibrium phase in which neither buying nor selling predominates. For $\Sigma > \Sigma_c$, two most probable values emerge that are symmetrical around zero demand, corresponding to excess demand and excess supply¹; we interpret this as an out-of-equilibrium phase in which the market behaviour is mainly buying for half of the time, and mainly selling for the other half.

We use the Trade and Quote database to analyse each and every transaction of the 116 most actively traded stocks in the two-year period 1994–95. We quantify demand by computing the volume imbalance, $\Omega(t)$, defined as the difference between the number of shares, Q_B , traded in buyer-initiated transactions and the number, Q_S , traded in seller-initiated transactions in a short time interval, Δt (refs 2, 3).

$$\Omega(t) \equiv Q_B - Q_S = \sum_{i=1}^N q_i a_i$$

where $i = 1, \dots, N$ labels each of the N transactions in the time interval Δt , q_i denotes the number of shares traded in transaction i , and $a_i = \pm 1$ denotes buyer-initiated and seller-initiated trades, respectively².

We also calculate, for the same sequence of intervals, the local noise intensity, $\Sigma(t) \equiv \langle |q_i a_i - \langle q_i a_i \rangle| \rangle$, where $\langle \dots \rangle$ denotes the local expectation value, computed from all transactions of that stock during the time interval Δt .

We find (Fig. 1a) that for small Σ , the conditional distribution, $P(\Omega|\Sigma)$, is single-peaked, displaying a maximum at zero demand, $\Omega = 0$. For Σ larger than a critical threshold, Σ_c , the behaviour of $P(\Omega|\Sigma)$ undergoes a qualitative change, becoming double-peaked with a pair of new maxima appearing at non-zero values of demand, $\Omega = \Omega_+$, and $\Omega = \Omega_-$, which are symmetrical around $\Omega = 0$.

Our findings for the financial-market problem are identical to what is known to occur in all phase-transition phenomena, wherein the behaviour of a system undergoes a qualitative change at a critical threshold, K_c , of some control parameter K . The change in behaviour at K_c can be quantified by an order parameter $\Psi(K)$, where $\Psi(K) = 0$ for $K < K_c$, and $\Psi(K) \neq 0$ for $K > K_c$.

For the financial-market problem, we find that the order parameter $\Psi = \Psi(\Sigma)$ is given by the values of the maxima of $\Omega_{\pm}$ of $P(\Omega)$. Figure 1b shows that the change in $\Psi(\Sigma)$ as a function of Σ is described by

$$\Psi(\Sigma) = \begin{cases} 0 & [\Sigma < \Sigma_c] \\ \Sigma - \Sigma_c & [\Sigma > \Sigma_c] \end{cases}$$

We interpret these two market phases as corresponding to the following two distinct conditions of the financial market.

First is the ' $\Sigma < \Sigma_c$ ' market phase, in which the distribution of demand, Ω , is single-peaked, with the most probable value being zero; we interpret this to be the market equilibrium phase, because the price of the stock is such that the probability of a transaction being buyer-initiated is equal to the probability of a transaction being seller-initiated⁴. In the equilibrium phase, there is statistically no net demand, and prices fluctuate around their 'equilibrium' values, suggesting that most of the trading is due to 'noise' traders who trade from misperceived information or for idiosyncratic reasons^{5–7}.

Second is the ' $\Sigma > \Sigma_c$ ' market phase, in which the distribution of demand is bimodal. We interpret this to be the out-of-equilibrium phase, because the price of the stock is such that there is an excess of either buyers or sellers and there is a non-zero net demand for stock. Thus, in the out-of-equilibrium phase, the prevalent 'equilibrium' price has changed, so the stock price is now being driven to the market's new evaluation of a fair value, which is consistent with the possibility that most of the trading arises from informed traders who possess superior information^{5–7}.

Our findings suggest that there is a link between the dynamics of a human system with many interacting participants (the financial market) and the ubiquitous phenomenon of phase transitions that occur in physical systems with many interacting units. Physical observables associated with phase transitions undergo large fluctuations that display power-law behaviour, so our results raise the possibility that volatile market movements and their empirically identified power-law behaviour are related to general aspects of phase transitions.

Vasiliki Plerou, Parameswaran

Gopikrishnan*, H. Eugene Stanley

Center for Polymer Studies and Department of Physics, Boston University, Boston, Massachusetts 02215, USA

e-mail: plerou@cgl.bu.edu

*Present address: Goldman Sachs, 10 Hanover Square, New York, New York 1005, USA

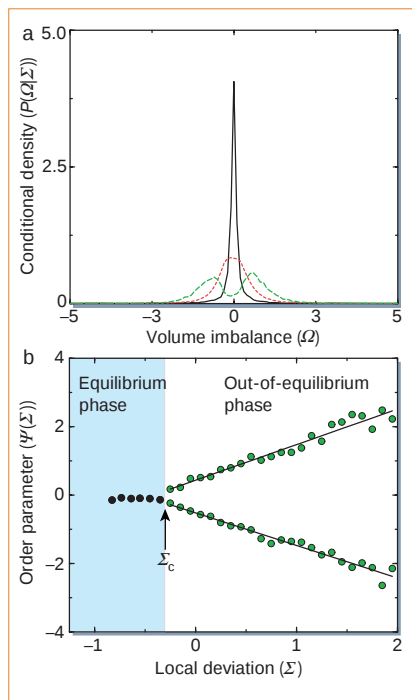


Figure 1 Empirical evidence supporting the existence of two distinct phases in a complex financial market. **a**, Conditional density, $P(\Omega|\Sigma)$, for varying Σ computed using data for all stocks. For each stock, Ω and Σ are normalized to zero mean and unit first centred moment. The distribution has a single peak for $\Sigma < \Sigma_c$ (solid line). For $\Sigma \approx \Sigma_c$ (dotted line), the distribution flattens near to the origin, and for $\Sigma > \Sigma_c$, $P(\Omega|\Sigma)$ displays two peaks (dashed line). **b**, Order parameter Ψ (positions of the maxima of the distribution $P(\Omega|\Sigma)$) as a function of Σ . For small Σ , $P(\Omega|\Sigma)$ displays a single maximum, whereas for large Σ , two maxima are present. To locate the extremes as accurately as possible, we compute all probability densities using the density estimator of ref. 8. Also shown (by shading) is a phase diagram representing the two distinct market phases. Here, $\Delta t = 15$ min; our results hold for Δt ranging from 15 min up to about half a day, beyond which our statistics are insufficient.

brief communications is intended to provide a forum for both brief, topical reports of general scientific interest and technical discussion of recently published material of particular interest to non-specialist readers. Priority will be given to contributions that have fewer than 500 words, 10 references and only one figure. Detailed guidelines are available on Nature's website (www.nature.com) or on request from nature@nature.com.

Continuing emissions of methyl chloroform from Europe

M. C. Krol*, J. Lelieveld†, D. E. Oram‡, G. A. Sturrock‡, S. A. Penkett‡, C. A. M. Brenninkmeijer†, V. Gros†, J. Williams† & H. A. Scheeren*†

* Institute for Marine and Atmospheric Research, Utrecht NL-3584 CC, The Netherlands

† Max Planck Institute for Chemistry, Mainz D-55128, Germany

‡ School for Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK

The consumption of methyl chloroform (1,1,1-trichloroethane), an industrial solvent, has been banned by the 1987 Montreal Protocol because of its ozone-depleting potential. During the 1990s, global emissions have decreased substantially and, since 1999, near-zero emissions have been estimated for Europe and the United States. Here we present measurements of methyl chloroform that are inconsistent with the assumption of small emissions. Using a tracer transport model, we estimate that European emissions were greater than 20 Gg in 2000. Although these emissions are not significant for stratospheric ozone depletion, they have important implications for estimates of global tropospheric hydroxyl radical (OH) concentrations, deduced from measurements of methyl chloroform. Ongoing emissions therefore cast doubt upon recent reports of a strong and unexpected negative trend in OH during the 1990s and a previously calculated higher OH abundance in the Southern Hemisphere compared to the Northern Hemisphere.

The principal uses of methyl chloroform (MCF) have been the degreasing of precision engineered components (31%) and cold cleaning (18%)¹. Its low toxicity compared to other non-flammable halogenated solvents favoured its use in many other applications such as dry cleaning, inks and coatings. The ultimate fate of MCF is evaporation into the atmosphere where its principal loss mechanism is oxidation by hydroxyl radicals (OH). The lifetime of MCF towards OH oxidation in the troposphere is relatively long (5–6 yr) and a significant MCF fraction is transported to the stratosphere where it releases chlorine through photolysis. For this reason, MCF was included in the Montreal Protocol (1987) and its amendments with a final phase-out in 1996 in developed countries, and 2015 in developing countries.

Compliance with the Montreal Protocol resulted in a rapid decrease of emissions and hence in the atmospheric MCF burden, beginning in the early 1990s (ref. 2). Currently, mean tropospheric concentrations are less than 40 pmol mol⁻¹, close to the levels that were observed before 1975, when the emissions were increasing rapidly. The highest atmospheric concentrations were measured around 1992, with background levels up to 160 pmol mol⁻¹. Until 1995, historical emissions of MCF were proficiently quantified because it is likely that all material produced finally ended up in the atmosphere. A detailed study using audited production data supplied by the chemical manufacturers showed that about 75% was released in the year of production and 25% in the following year¹. In the early 1990s some stockpiling took place and the total banked quantities were estimated to be around 5 Gg for 1997–2000 (ref. 3). From 1995 onwards the emission estimates have been based on the non-audited consumption and production data supplied by the parties under the Montreal Protocol^{3,4}.

Long-term measurements of MCF in the atmosphere have been combined with the emission estimates provided by industry and inverse modelling procedures to derive the mean MCF lifetime in the troposphere^{5–10} and thereby the average atmospheric OH concentration. The temporal evolution of the average atmospheric OH concentration has been calculated using the 1978–2000 data obtained from the ALE/GAGE/AGAGE programme at remote stations^{6–8}. From these analyses, using the most recent estimates of MCF emissions, it has been inferred that global OH levels increased by 5% or more during the 1980s (refs 8–10). However, the recent decrease in the atmospheric MCF concentrations is less

rapid than was anticipated from the phase-out of MCF production and use in the 1990s, assuming constant OH levels. These observations have been explained by a substantial and unexpected drop in tropospheric OH since 1990 (ref. 8).

Prinn *et al.*⁸ attributed the inferred tropospheric mean OH reduction mainly to changes in the Northern Hemisphere, for which they found that OH levels are on average 14% lower than in the Southern Hemisphere. Higher OH levels south of the intertropical convergence zone (ITCZ) have also been inferred from NOAA flask samples combined with a budget analysis¹¹. The asymmetry in OH distribution inferred from the MCF measurement programmes is a contentious result, with some global chemistry/transport models predicting more OH in the Northern Hemisphere compared to the Southern Hemisphere^{12–14}. These

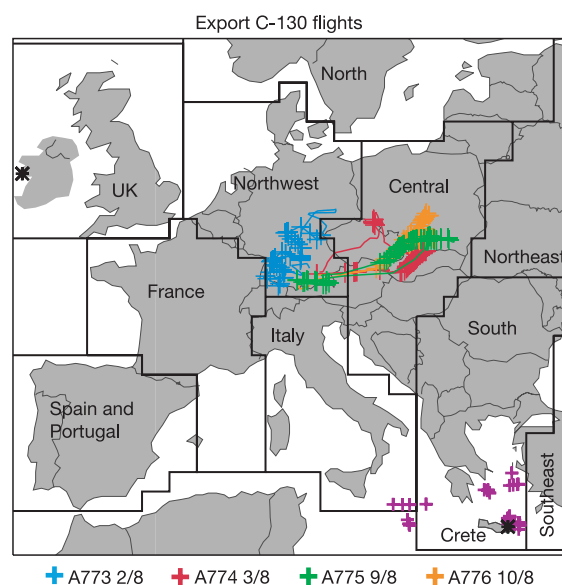


Figure 1 Flight tracks of the British C-130 aircraft. Coloured crosses indicate the locations at which the air samples were taken. The asterisk on Ireland shows the location of Mace Head. In sensitivity simulations, 4 Gg MCF yr⁻¹ has been added to ten regions within Europe (northeastern and southeastern Europe extend further eastwards). The asterisk on Crete and the purple crosses around Crete refer to the MINOS campaign.

model results may be flawed by wrongly simulated OH precursor distributions. Calculated hemispheric OH concentrations become about equal if observed OH precursor fields are used¹⁵, although appreciable uncertainties remain.

All methods that use MCF to estimate OH levels, trends, and hemispheric distributions critically rely on the accuracy of the MCF emission estimates and the accuracy of the MCF absolute calibration. The observed decline in atmospheric concentrations at remote sites is indeed consistent with a significant drop in emissions. Not only have the measured concentrations declined rapidly, but the observed atmospheric variability has also decreased dramatically, indicating that emissions no longer occur, at least not in the regions where the measurement stations are located.

Evidence for methyl chloroform emissions

We present MCF measurements taken in the free troposphere and the boundary layer over central Europe during the EXPORT (European export of precursors and ozone by long-range transport) experiment involving the UK Met Office C-130 aircraft (see Methods). The MCF measurements have been interpreted using tracer-transport model simulations (see 'Model description' section). The accuracy of the modelled transport has been verified using CO observations that were taken during the EXPORT flights and at Mace Head (Ireland, see Fig. 1)². To locate potential MCF

emission regions, several model 'experiments' have been carried out that include additional European emissions. In each experiment, the European emissions were enhanced by 4 Gg yr^{-1} over one of the regions depicted in Fig. 1. In an additional simulation, the emissions were enhanced with 4 Gg yr^{-1} over eastern North America.

Figure 2 shows the simulated MCF (Fig. 2a–d) and CO (Fig. 2e–h) concentrations as a function of height during four flights of the EXPORT measurement campaign. The reference simulation (1 Gg yr^{-1} MCF emissions over Europe, red line in Fig. 2a–d) shows negligible concentration variations. This indicates that the much higher and more variable concentrations observed are the result of unexpected MCF emissions. On the basis of the sensitivity simulations described below, possible source regions have been located in Spain and Portugal ($\sim 6 \text{ Gg yr}^{-1}$), Italy ($\sim 9 \text{ Gg yr}^{-1}$), central Europe ($\sim 3 \text{ Gg yr}^{-1}$), northwestern Europe ($\sim 3 \text{ Gg yr}^{-1}$), and France ($\sim 2 \text{ Gg yr}^{-1}$). Although the estimated emissions and their geographic locations (broadly outlined in Fig. 1) are rather uncertain, the simulated concentrations (with 24 Gg yr^{-1} emissions over Europe; blue line in Fig. 2a–d) are in much better agreement with the measurements.

The 2 August 2000 flight (A773) over Germany (in Fig. 2a) shows increased concentrations in the free troposphere around 4 km altitude. From the sensitivity simulations, it can be inferred that the probable source of these high concentrations is in Spain (black line in Fig. 2a), although contributions from France cannot be ruled out. A trajectory analysis indicates that stagnant air over northern Spain was lifted over the Pyrenees by frontal thunderstorms before being advected in the free troposphere to Germany.

Similar observations were made on 3 August 2000 (A774) when the front moved towards the Alps. The model simulations and trajectory analysis indicate that, in this case, the source of the increased concentrations is located in Italy (black line in Fig. 2b). Again, stagnant and polluted air parcels were lifted over the Alps, followed by advection in the free troposphere to central Europe. The last part of the flight was at altitudes below 2,000 m and indicates possible sources over northwestern Europe, France and central Europe.

The 9 August 2000 flight (A775) was performed mostly in the boundary layer over central Europe. The increased MCF concentrations observed are probably caused by emissions in central Europe (black line in Fig. 2c) and northwestern Europe. Finally, the boundary layer concentrations during the 10 August 2000 flight (A776) are most probably explained by emissions from central Europe (black line in Fig. 2d) and from northwestern Europe.

Although the simulations with enhanced European emissions are closer to the observations, the agreement is qualitative and individual peaks are not resolved. This is only to be expected, because the exact locations of the emissions are unknown. The model resolution is also limited and model errors that are associated with, for example, vertical transport may be appreciable. The simulation of the CO distribution during the EXPORT flights (Fig. 2e–h) nevertheless shows that the model reproduces the observed CO enhancements in the free troposphere and in the boundary layer. Only the remarkably low CO concentrations during flight A776 were not resolved by the model. For the rest, the observed vertical CO distribution of all flights is simulated adequately, which indicates that the modelled vertical transport is realistic. The simulations furthermore indicate that European MCF emissions of more than 20 Gg yr^{-1} are needed to explain the large range of observed concentrations. With no or very small additional emissions, the calculated MCF variability is absent or unrealistically small. Hence there is little doubt that the emissions are substantially higher than the $<1.0 \text{ Gg yr}^{-1}$ estimate for Europe³.

Ground-based measurements

Routine monitoring of MCF in Europe has historically been limited to the background station Mace Head in Ireland (Fig. 1). To

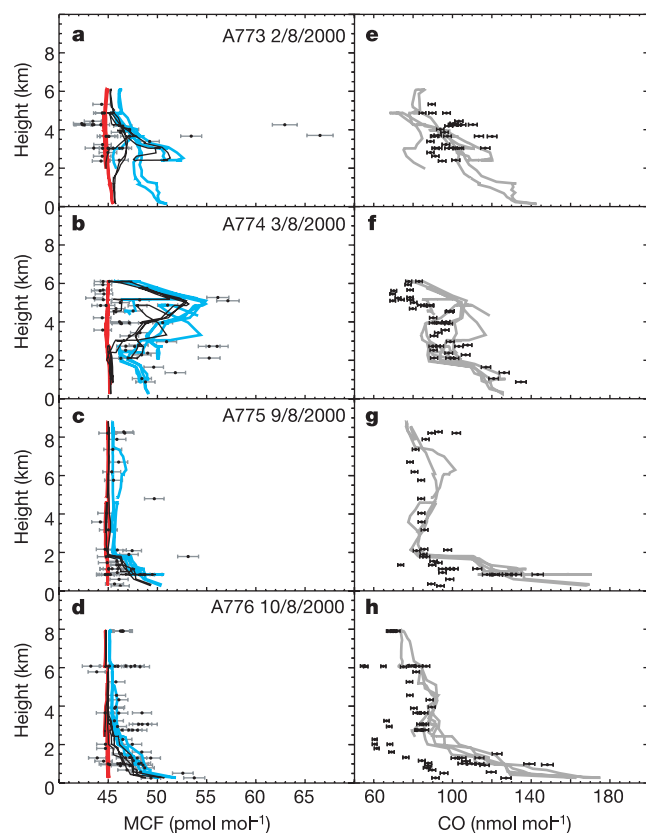


Figure 2 Analysis of the EXPORT flights as a function of height. The number and date of each flight is given. The black dots indicate the measured methyl chloroform (MCF) (pmol mol^{-1}) and CO (nmol mol^{-1}) concentrations and the bars represent the measurement errors. **a–d.** The red lines show the modelled MCF concentrations along the flight tracks in the reference simulation (1 Gg yr^{-1} European emissions). The blue lines represent the concentrations in a simulation with enhanced emissions in the regions Spain and Portugal (6 Gg yr^{-1}), Italy (9 Gg yr^{-1}), central Europe (3 Gg yr^{-1}), northwestern Europe (3 Gg yr^{-1}) and France (2 Gg yr^{-1}). The thin black lines refer to enhanced emissions only in Spain and Portugal (A773), Italy (A774), and central Europe (A775 and A776). **e–h.** The grey lines show the simulated CO concentrations along the EXPORT flight tracks corresponding to **a–d**.

investigate the extent to which the additional emissions in Europe would influence the concentrations at Mace Head during the summer of 2000, Fig. 3a and b compare the model calculated CO and MCF time series to the observations from mid-May to late August 2000. The reference MCF simulation (1 Gg yr^{-1} emissions over Europe, red lines in Fig. 3b and c) appears quite consistent with the observations. The observed and modelled MCF fluctuations are associated with meteorological variability. Air masses originating from the subtropics (southerly winds) are generally more strongly depleted in MCF, which is caused by the higher OH levels in these regions. The model captures these fluctuations quite well. These findings are supported by the CO simulation, which shows generally good agreement with the CO variability observed at Mace Head.

The temporal MCF decline is largely caused by OH oxidation, which reaches a maximum in summer. The calculated decline in the simulation with enhanced MCF emissions (24 Gg yr^{-1} emissions over Europe, blue lines) is about 13% smaller than in the reference simulation. Owing to the prevailing westerly winds, European MCF emissions normally reach Mace Head only after hemispheric transport, so that European emissions cannot be detected directly at Mace Head. Some pollution events are evident from the CO measurements and simulation. To facilitate comparison with the MCF sensitivity simulations (Fig. 3d), Fig. 3c shows the MCF series with the trends removed.

The perturbation simulation with the additional 23 Gg yr^{-1} of MCF emissions from Europe (blue line in Fig. 3c) reveals that some observed concentration enhancements are now captured better by the model. On some days, however, the modelled enhancements are not measured at Mace Head. This is not unexpected, because the emissions have been applied over relatively large regions, weighted with the population density. A number of point sources would probably be more realistic, but the unknown nature of the MCF sources precludes such an approach.

The sensitivity calculations shown in Fig. 3d indicate that during the summer of 2000 the following occurred: (1) air masses from the regions labelled Italy and central Europe (and southern, north-eastern and southeastern Europe—not shown) reach Mace Head only in hemispherically diluted form; (2) substantial eastern USA sources ($>4 \text{ Gg yr}^{-1}$) do not seem to be consistent with the observations at Mace Head; and (3) some remaining discrepancies between model and measurements can be explained by small emissions from the regions labelled UK and North in Fig. 1. Most importantly, the additional European MCF emissions required to explain the EXPORT measurements are not in conflict with the Mace Head observations.

Previous attempts to estimate European source strengths from the Mace Head MCF observations indicated that MCF emissions dropped from about 30 Gg yr^{-1} in 1995 to about 2 Gg yr^{-1} in 1998 (ref. 16). We note, however, that this study¹⁶ confirms that sources from Italy, Spain, central and eastern Europe do not significantly contribute to the measured concentrations at Mace Head.

Additional evidence

The unexpected large MCF variability, as observed during EXPORT, is also confirmed by MCF measurements by aircraft over the eastern Mediterranean and at a background station on Crete, performed during the Mediterranean intensive oxidant study (MINOS)¹⁷ in August 2001 (see Figs 1 and 4). Boundary layer concentrations well above the background were regularly observed in air masses transported from southern and eastern Europe. Back-trajectory analysis suggests that the peak value of about 50 pmol mol^{-1} observed north of Crete (Fig. 4a) was caused by eastern European sources with a possible contribution from Greece. This peak coincided with the second peak measured at the surface (13–14 August, Fig. 4b).

Figure 4c shows MCF measurements from samples collected

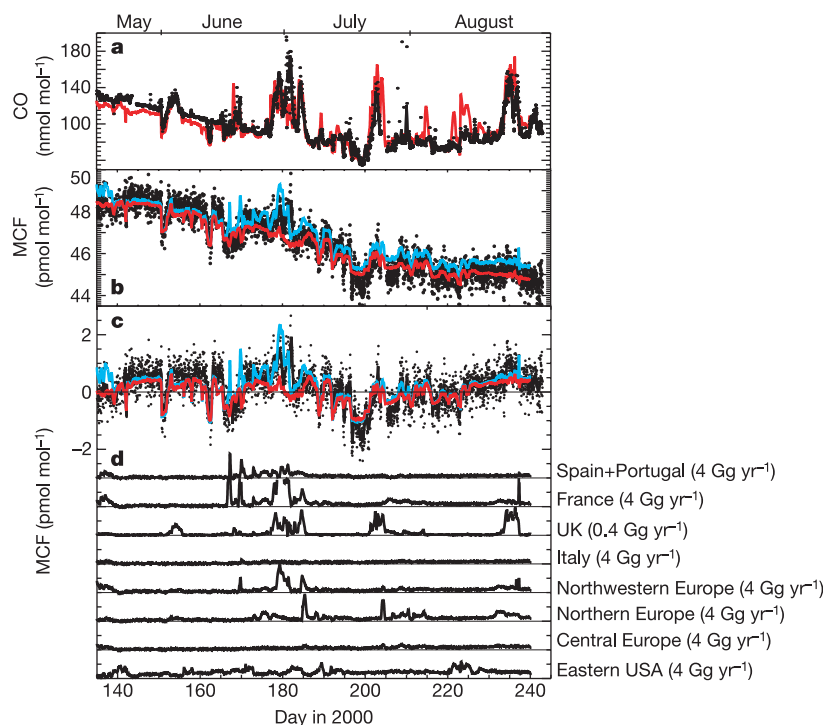


Figure 3 MCF and CO simulations compared to Mace Head observations. **a**, CO measurements (dots, the black line represents 6-hourly means) and corresponding model calculations (red). **b**, MCF measurements (dots; the black lines represent 6-hourly means) and model calculations (red and blue). **c**, The data from **b** normalized to zero and the trend removed (about 3 pmol mol^{-1} over this period). The red lines represent the reference simulation with 1 Gg yr^{-1} European MCF emissions, and the blue lines the simulation with

23 Gg yr^{-1} additional emissions (see legend of Fig. 2). **d**, Calculated concentration perturbations (relative to the reference simulation and on the same scale) with 4 Gg yr^{-1} additional emissions in the regions defined in Fig. 1 and in the eastern USA. Considering the large effect on the MCF concentration at Mace Head, the additional emissions in the UK region have been assumed to be 0.4 Gg yr^{-1} only.

during upper tropospheric aircraft measurements in the CARIBIC project (civil aircraft for the regular investigation of the atmosphere based on an instrument container)¹⁸. The MCF trend between 1998 and 2001 is consistent with that observed from background stations at the surface^{8,11}. The observed concentrations at flight altitude (200–300 hPa) are rather variable. The highest concentration of nearly 75 pmol mol⁻¹ was observed in August 1999 near India, associated with rapid upward transport of polluted air in monsoon convection. These MCF enhancements in the upper troposphere indicate that significant emissions may also occur outside Europe.

Possible sources of methyl chloroform

The MCF concentrations measured during EXPORT are strongly correlated with tetrachloromethane (CCl₄) observations ($R^2 = 0.6$). Production of CCl₄, largely used as an intermediate in the production of chlorofluorocarbons (CFCs), is also regulated under the Montreal Protocol, and the high concentration fluctuations point to considerable emissions of this compound as well. Moreover, the high correlation suggests that both compounds have a very similar source signature. It is unlikely that this result is associated with similar loss processes in the atmosphere because the removal of CCl₄ is controlled by stratospheric photolysis, whereas MCF is predominantly removed by tropospheric OH.

A natural origin for the MCF emissions can also be excluded, considering the near absence of MCF in pre-industrial air retrieved from polar firn¹⁹. If the consumption data of the European countries are reliable, current use of MCF in the traditional applications can be ruled out. However, the MCF surface concentrations measured during MINOS were highly correlated with CFC-113 observations ($R^2 = 0.5$). Because the applications of CFC-113 are similar to those of MCF, this may indicate active use of both compounds, at

least in southern and eastern parts of Europe. MCF is produced for non-dispersive uses, but widespread losses during production are not expected. Alternatively, the banking of MCF in consumer goods may have been underestimated. MCF in coatings, inks and textiles represents a considerable fraction of its total use before 1996 (ref. 1) and the distribution and use of these products is rather uncertain. They may have ended up in waste and landfills from which MCF slowly diffuses into the atmosphere. The fact that MCF sources are probably located in southern and central Europe, where waste dumping is more common than waste burning, gives some credence to this explanation. Moreover, diffusion from landfills would be more efficient in summer owing to the higher temperatures. An alternative explanation involves (chemical) waste burning of MCF-containing material. There is some evidence that this compound is still present in the exhaust gases after high-temperature burning of MCF-containing mixtures²⁰.

Implications for OH trends

The unexpected remaining MCF emissions are small compared to the amounts that were emitted in the early 1990s. Given the low ozone depletion potential of MCF compared to CFCs⁴, the effect of these emissions on stratospheric ozone depletion will be relatively insignificant. However, the continuing emissions strongly influence analyses of global OH concentrations as well as the inferred hemispheric OH ratio. Montzka *et al.*¹¹ performed a hemispheric budget analysis based on the MCF burden, sources and sinks. They showed that 1998–99 emissions of about 40 Gg yr⁻¹ in the Northern Hemisphere would lead to a south-to-north hemispheric OH ratio of about 1. On the other hand, if MCF emissions in the Northern Hemisphere were very small, about 15% more OH in the Southern Hemisphere compared to the Northern Hemisphere would explain the MCF measurements. Our analysis, however, suggests that in Europe alone, emissions of more than 20 Gg yr⁻¹ would be consistent with the MCF measurements reported here, thus rendering a strong interhemispheric OH asymmetry, with higher OH in the Southern Hemisphere, unlikely.

A similar argument applies to MCF-derived OH trends since the late 1970s. Prinn *et al.*⁸ estimated the MCF sources that would correspond to a time-invariant global tropospheric OH abundance. For the years 1996–2000, the required additional MCF sources would have to be 17, 41, 29, 16 and 14 Gg yr⁻¹, respectively. Again, even by considering the European MCF emissions alone, a strong negative OH trend during the 1990s seems unlikely. Finally, if the emissions did not cease as expected in recent years, and the gross MCF production has been estimated correctly, more MCF should have been attributed to the ‘slow use’ category. If the additional emissions during the 1990s would in fact represent delayed release into the atmosphere, not only would the presumed negative trend in this period vanish, but the positive OH trend in the preceding decade would also be much smaller than previously inferred^{8,9}. We estimated that the large OH anomaly around 1990 that is calculated with the emission estimates used by ref. 8 disappears if MCF emissions of about 65 Gg were delayed from the early to the late 1990s¹⁰. The implication is that tropospheric OH has been relatively constant in time over the past decades, being more consistent with observed methane growth rates and with atmospheric chemistry modelling^{21,22}.

Definite conclusions about the global OH distribution and trend cannot be drawn before the emissions and distribution of MCF are better quantified. Additional measurements are required to locate possible sources. □

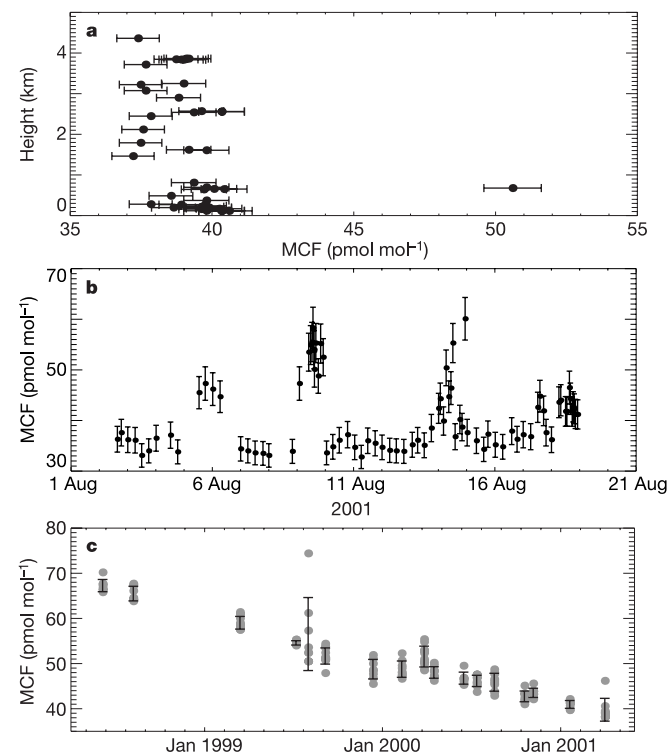


Figure 4 Methyl chloroform measurements performed in the MINOS and CARIBIC projects, showing substantial variability up to 2001. **a**, The black dots represent MCF measurements (and uncertainty bars) from 15 flights over the eastern Mediterranean Sea in August 2001 (see purple crosses in Fig. 1). **b**, Ground-based MCF measurements from Crete (35° N, 26° E). **c**, MCF measurements (~120 grey dots), including flight mean and 1σ standard deviation, from samples taken in the upper troposphere during 16 flights between the Maldives/Sri Lanka and Germany.

Methods

Experimental method

During EXPORT, whole-air samples (not dried) were collected in 3-litre, silica-lined, stainless steel canisters (Restek Corporation) pressurized to 2–3 atm using a stainless steel, double-headed bellows pump (Metal Bellows) to draw air from the aircraft air sample inlet. The samples were analysed for a large number of halocarbons, including MCF, as

soon as possible (1–20 days after each flight) using a fully automated gas chromatograph/mass spectrometer (Agilent 6890/5973 GC-MSD) in select ion monitoring mode. An 800-ml sample from each flask was processed using an Etech 7100 pre-concentrator system (Etech Instruments Inc.) by sampling at 100 ml min⁻¹ onto an 1/8-inch (outer diameter) stainless steel trap (trap 1) packed with glass beads and held at -150 °C. To facilitate the removal of water and carbon dioxide from the enriched sample, the contents of trap 1 were swept onto a second trap (same size and material as trap 1) packed with Tenax at -40 °C. Any water in the original sample remained on trap 1, whereas CO₂ passed through trap 2 without significant retention. A final trapping stage involved cryo-focusing on fused silica-lined stainless steel tubing (1/32-inch) before transfer onto the gas chromatography column. A DB-5 capillary column (J&W Scientific, 105 m × 0.32 mm inner diameter, 1.5 µm film thickness) with helium carrier gas (UHP) was used for temperature-programmed separation before detection. Analytical precision was typically ± 2%. Samples of MCF-free, zero air were analysed to ensure there was no contamination from either the analytical system or the aircraft sampling manifold. All air samples were analysed against one of a series of secondary whole-air samples stored in 3.2-litre, electropolished, stainless steel flasks. These working standards were calibrated against a high-pressure whole-air sample stored in an Acuflex-treated, aluminium cylinder supplied by NOAA-Climate Monitoring and Diagnostics Laboratory (CMDL). The MCF measurements are reported on the most recent NOAA-CMDL calibration scale (CMDL-2002). Overall uncertainty in the measurements, which includes uncertainties in the NOAA standard, propagation of working standards and experimental precision, was approximately 5%. The CARIBIC MCF measurements were obtained using a combination of the system described above and a second GC-MS technique described elsewhere^{23,24}.

The CO measurements were made with a fast-response, vacuum-ultraviolet resonance fluorescence instrument²⁵. It has a 1-s time resolution, a detection limit of 3 nmol mol⁻¹ and a precision of ± 1.5 nmol mol⁻¹ at an atmospheric mixing ratio of 100 nmol mol⁻¹. In-flight calibrations were performed approximately every 30 min using a British Oxygen Company (BOC) alpha standard (± 1%). The high-frequency CO measurements were aggregated to reproduce the sampling intervals of the MCF measurements.

The MINOS flight canisters were analysed using a gas chromatograph equipped with an electron capture detector at Utrecht University²⁶. The canisters collected at the ground station were analysed at the Max Planck Institute, Mainz, with a GC-MS (Agilent 6890/5973) similar to that described above.

Model description

The MCF distribution over Europe at the time of the EXPORT experiment (August 2000) was simulated using the global Tracer Model, version 5 (TM5), that zooms in over Europe. The global and regional simulations were coupled by means of two-way nesting^{27,28}. The European simulations were initialized by monthly averaged MCF distributions. The initial MCF field was taken from a long-term global simulation (1951–2000) with a coarse-grid model version. Monthly averaged OH fields were used to oxidize MCF. These fields have been scaled such that the observed decline in MCF in the summer of 2000 was reproduced¹⁰. The initial MCF concentration field at the start of the simulations (1 May 2000) was scaled to match the Mace Head observations. The simulations accounted for removal by OH oxidation, oceanic uptake and MCF destruction by photolysis in the stratosphere²⁹.

Carbon monoxide (CO) distributions have been simulated to verify the modelled transport. Initial fields were taken from a full chemistry simulation²² and CO was oxidized by the same OH fields that were used for MCF. Anthropogenic CO emissions are based on the EDGAR3.2 database³⁰ for the year 1995. Emissions from fossil fuel amount to 296 Tg CO yr⁻¹; from biofuel to 235 Tg CO yr⁻¹; from biomass burning to 256 Tg CO yr⁻¹; and from temperate fires to 58 Tg CO yr⁻¹. The highly uncertain natural emissions from vegetation, soils and oceans are taken as 115 Tg CO yr⁻¹. Apart from surface emissions, an additional CO source from methane oxidation has been accounted for, assuming constant methane levels of 1.8 µmol mol⁻¹.

Transport is performed using 6-hourly fields from the European Centre for Medium range Weather Forecasts (ECMWF). All physical parameterizations in the zoom model are identical to the global model as used in previous studies^{31,32} and the boundary layer representation is based on 3-hourly surface fields to resolve better the diurnal cycle of near-surface mixing³³. The modelled concentrations are evaluated every 15 min, which corresponds to the integration time step within the European model domain. The global resolution of the model is 6° latitude by 9° longitude, and the European region is resolved at a higher horizontal resolution of 1° × 1°. In the vertical direction, 25 of the 60 layers used by the parent ECMWF model are retained, mostly in the boundary layer and the free troposphere. Concentrations are interpolated to the C-130 aircraft EXPORT flight tracks. In addition, the MCF and CO concentrations are evaluated at the AGAGE station Mace Head in western Ireland (53° N, 10° W).

From 1980 onward, the estimated annual global MCF emissions were divided over seven distinct regions^{1,3}. Within these regions emissions were distributed according to the population density. The global total MCF source applied in the reference model simulation for the year 2000 amounts to 19.7 Gg³, mostly emitted outside North America and Europe. European and North American emissions for the year 2000 were assumed to be 1 and 5 Gg, respectively. These amounts are already somewhat higher than the most recent emission compilation, which reports that consumption in Europe and North America had already dropped to zero in 1999 (ref. 3).

Received 13 September; accepted 15 November 2002; doi:10.1038/nature01311.

1. Midgley, P. M. & McCulloch, A. The production and global distribution of emissions to the atmosphere of 1,1,1-trichloroethane (methyl chloroform). *Atmos. Environ.* **29**, 1601–1608 (1995).
2. Prinn, R. G. *et al.* A history of chemically and radiatively important gases in air deduced from ALE/GAGE/AGAGE. *J. Geophys. Res.* **105**, 17751–17792 (2000).
3. McCulloch, A. & Midgley, P. M. The history of methyl chloroform emissions: 1951–2000. *Atmos.*

Environ. **35**, 5311–5319 (2001).

4. UNEP (United Nations Environment Programme). *Production and Consumption of Ozone-depleting Substances under the Montreal Protocol: 1986–2000* (UNEP, Nairobi, Kenya, 2002); available at <http://www.unep.org/ozone/15-year-data-report.pdf>.
5. Singh, H. B. Preliminary estimation of average tropospheric HO concentrations in the Northern and Southern Hemisphere. *Geophys. Res. Lett.* **4**, 453–456 (1977).
6. Prinn, R. G. *et al.* Atmospheric trends in methylchloroform and the global average for the hydroxyl radical. *Science* **238**, 935–950 (1987).
7. Prinn, R. G. *et al.* Atmospheric trends and lifetime of CH₂Cl₂ and global OH concentrations. *Science* **269**, 187–192 (1995).
8. Prinn, R. G. *et al.* Evidence for substantial variations of atmospheric hydroxyl radicals in the past two decades. *Science* **229**, 1882–1888 (2001).
9. Krol, M., van Leeuwen, P. J. & Lelieveld, J. Global OH trend inferred from methylchloroform measurements. *J. Geophys. Res.* **103**, 10697–10711 (1998).
10. Krol, M. & Lelieveld, J. Can the variability in tropospheric OH be deduced from measurements of 1,1,1-trichloroethane (methyl chloroform)? *J. Geophys. Res.* (in the press).
11. Montzka, S. A. *et al.* New observational constraints for atmospheric hydroxyl on global and hemispheric scales. *Science* **288**, 500–503 (2000).
12. Hauglustaine, D. A. *et al.* MOZART, a global chemical transport model for ozone and related chemical tracers 2. Model results and evaluation. *J. Geophys. Res.* **103**, 28291–28335 (1998).
13. Houweling, S., Dentener, F. & Lelieveld, J. The impact of nonmethane hydrocarbon compounds on tropospheric photochemistry. *J. Geophys. Res.* **103**, 10673–10696 (1998).
14. Wang, Y. & Jacob, J. D. Anthropogenic forcing on tropospheric ozone and OH since preindustrial times. *J. Geophys. Res.* **103**, 31123–31135 (1998).
15. Spivakovskiy, J. *et al.* Three-dimensional climatological distribution of tropospheric OH: Update and evaluation. *J. Geophys. Res.* **105**, 8931–8980 (2000).
16. Ryall, D. B., Derwent, R. G., Manning, A. J., Simmonds, P. G. & O'Doherty, S. Estimating source regions of European emissions of trace gases from observations at Mace Head. *Atmos. Environ.* **35**, 2507–2523 (2001).
17. Lelieveld, J. *et al.* Global air pollution crossroads over the Mediterranean. *Science* **298**, 794–799 (2002).
18. Brenninkmeijer, C. A. M. *et al.* CARIBIC - Civil aircraft for global measurement of trace gases and aerosols in the tropopause region. *J. Atmos. Ocean. Technol.* **16**, 1373–1383 (1999).
19. Butler, J. H. *et al.* A record of atmospheric halocarbons during the twentieth century from polar firn air. *Nature* **399**, 749–755 (1999).
20. Dziura, E. L., Molseed, A. C. & Kramlich, J. C. Thermal destruction behavior of selected waste compounds under short-time, high quench rate conditions. *Environ. Eng. Sci.* **14**, 33–42 (1997).
21. Lelieveld, J., Peters, W., Dentener, F. J. & Krol, M. Stability of tropospheric hydroxyl chemistry. *J. Geophys. Res.* **107**, 10.1029/2002JD002272 (2002).
22. Dentener, F., van Weele, M., Krol, M., Houweling, S. & van Velthoven, P. Trends and inter-annual variability of methane emissions derived from 1979–1993 global CTM simulations. *Atmos. Chem. Phys. Discuss.* **2**, 249–287 (2002).
23. Oram, D. E., Reeves, C. E., Penkett, S. A. & Fraser, P. J. Measurements of HCFC-142b and HCFC-141b in the Cape Grim air archive: 1978–1993. *Geophys. Res. Lett.* **22**, 2741–2744 (1995).
24. Sturges, W. T. *et al.* Methyl bromide, other brominated methanes, and methyl iodide in polar firn air. *J. Geophys. Res.* **106**, 1595–1606 (2001).
25. Gerbig, C., Schmitgen, S., Kley, D. & Volz-Thomas, A. An improved fast-response vacuum-UV resonance fluorescence CO instrument. *J. Geophys. Res.* **104**, 1699–1704 (1999).
26. Scheeren, H. A., Lelieveld, J., de Gouw, J. A., van der Veen, C. & Fischer, H. Methyl chloride and other chlorocarbons in polluted air during INDOEX. *J. Geophys. Res.* **107**, 10.1029/2001JD001121 (2002).
27. Berkvens, P. J. F., Botchev, M. A., Lioen, W. M. & Verwer, J. G. A zooming technique for wind transport of air pollution. Report MAS-R9921 (CWI, Amsterdam, 1999).
28. Krol, M. C., Peters, W., Berkvens, P. J. F. & Botchev, M. A. A new algorithm for two-way nesting in global models: principles and applications. *Proc. 2nd Int. Conf. Air Pollution Modelling and Simulation, April 9–12, 2001, Champs-sur-Marne* (ed. Sportisse, B.) 225–235 (Springer, Berlin, Heidelberg and New York, 2002).
29. Krol, M., van Leeuwen, P. J. & Lelieveld, J. Comment on “Global OH trend inferred from methylchloroform measurements” by Maarten Krol *et al.* - Reply. *J. Geophys. Res.* **106**, 23159–23164 (2001).
30. Olivier, J. G. J. & Berdowski, J. J. M. In *The Climate System* (eds Berdowski, J., Guicherit, R. & Heij, B. J.) 33–78 (Balkema/Swets & Zeitlinger, Lisse, 2001).
31. Lelieveld, J. & Dentener, F. J. What controls tropospheric ozone? *J. Geophys. Res.* **105**, 3531–3551 (2000).
32. Peters, W., Krol, M., Dentener, F. & Lelieveld, J. Identification of an El Niño-Southern Oscillation signal in a multiyear global simulation of tropospheric ozone. *J. Geophys. Res.* **106**, 10389–10402 (2001).
33. Jeuken, A., Veeckind, J. P., Dentener, F., Metzger, S. & Gonzalez, C. R. Simulation of the aerosol optical depth over Europe for August 1997 and a comparison with observations. *J. Geophys. Res.* **106**, 28295–28311 (2001).

Acknowledgements Assistance from E. Atlas and S. Montzka in instrumental design and calibrations is gratefully acknowledged. R.G. Prinn and colleagues provided the AGAGE MCF and CO data at Mace Head. We thank F. J. Dentener for comments and help with the implementation of the EDGAR 3.2 CO emissions, which were provided by J. Olivier. We also thank C. v.d. Veen and R. Hofmann for experimental assistance. EXPORT CO data were provided by H. Barjat and J. Kent of the UK Met Office using an instrument on loan from the Institut für Chemie und Dynamik der Geosphäre Forschungszentrum Jülich. EXPORT was funded by the UK Natural Environment Research Council (NERC) with additional funds from the UK Department of Environment, Food and Rural Affairs (DEFRA) and the UK Met Office. M.K. is supported by the Space Research Organisation Netherlands (SRON). V.G. is supported by a scientific cooperation between MPG and CNRS.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.K. (e-mail: krol@phys.uu.nl).

Modelling vaccination strategies against foot-and-mouth disease

M. J. Keeling*, M. E. J. Woolhouse†, R. M. May‡, G. Davies§ & B. T. Grenfell||

* Department of Biological Science & Mathematics Institute, University of Warwick, Gibbet Hill Road, Coventry, CV4 7AL, UK

† Centre for Infectious Diseases, College of Medicine and Veterinary Medicine, University of Edinburgh, Easter Bush, Roslin, Midlothian, EH25 9RG, UK

‡ Department of Zoology, University of Oxford, Oxford, OX1 3PS, UK

§ Zinnia, Kettlewell Hill, Woking, Surrey, GU21 4JJ, UK

|| Department of Zoology, University of Cambridge, Downing Street, Cambridge, CB2 3EJ, UK

Vaccination has proved a powerful defence against a range of infectious diseases of humans and animals. However, its potential to control major epidemics of foot-and-mouth disease (FMD) in livestock is contentious. Using an individual farm-based model, we consider either national prophylactic vaccination campaigns in advance of an outbreak, or combinations of reactive vaccination and culling strategies during an epidemic. Consistent with standard epidemiological theory, mass prophylactic vaccination could reduce greatly the potential for a major epidemic, while the targeting of high-risk farms increases efficiency. Given sufficient resources and preparation, a combination of reactive vaccination and culling might control ongoing epidemics. We also explore a reactive strategy, 'predictive' vaccination, which targets key spatial transmission loci and can reduce markedly the long tail that characterizes many FMD epidemics. These analyses have broader implications for the control of human and livestock infectious diseases in heterogeneous spatial landscapes.

The epidemic of FMD in 2001 exerted a major toll on Great Britain's livestock industry and highlighted a series of important questions about the management of infectious diseases in animal populations¹. Detailed records of this epidemic² have allowed the development of epidemiological models that capture the main features of epidemic spread in space and time, and the impact of control strategies^{3–7}.

After the introduction of livestock movement restrictions on 23 February 2001, the epidemic showed intense local stochastic spread, together with rarer long-distance jumps^{2,3,7}. This generated a severe initial epidemic in localized regions, followed by a long, irregular epidemic tail. Control measures used during the 2001 FMD epidemic were movement restrictions, biosecurity and culling, coupled with surveillance. Different models all indicated that prompt culling of infected premises (IPs) and at-risk farms (dangerous contacts, DCs, and contiguous premises, CPs) was necessary to control the disease because of the intense local spread^{3,5–7}. Although reactive vaccination of livestock during the epidemic was proposed, various biological and logistical problems confounded its likely effectiveness (see Supplementary Information). Here, we consider the potential impact of both reactive and prophylactic vaccination^{8,9} on future FMD epidemics, in terms of expected epidemic size and duration, for a variety of models. We concentrate on cattle vaccination, but also compare this default with vaccinating other livestock species. The fine spatial grain of the British data also allows us to explore two general epidemiological questions. First, how well does the performance of vaccination in this complex spatial epidemic match the predictions of standard theory⁸? Second, can we use our detailed spatial model to generate predictive vaccination strategies, which optimize disease control by targeting key 'nodes' in the spatial transmission network?

To model the dynamics of vaccination, we need to allow for the biological limitations of current FMD vaccine formulations. Major determinants of the effectiveness of prophylactic vaccination are uptake (the proportion of animals vaccinated) and efficacy (the proportion of vaccinated animals that are protected)^{10–12}. Efficacy (typically no more than 90–95%¹³) can be confounded by antigenic variation between FMD virus strains, requiring some prior assessment of which strains present the greatest threat. Reactive vaccination

evades the problem of strain uncertainty but suffers from a significant delay (around 4 days for high-potency vaccines and 10 days for standard formulations¹⁴) before vaccinated animals are protected⁷.

To focus the analysis, simulations are based on the highly disseminated 2001 epidemic in Great Britain as a pessimistic model; however, we interpret these results in terms of the generic performance of future vaccination campaigns, where appropriate stocks of highly effective vaccines and strategies for delivering them are in place before the start of the epidemic. We address these issues using a relatively parsimonious individual-based stochastic model (see Methods), which accurately captures the overall spatio-temporal pattern of the 2001 epidemic in Great Britain⁷ (see Supplementary Information for a discussion of parameter sensitivity and model assumptions).

Prophylactic vaccination

Figure 1 reports the impact of prophylactic vaccination under a variety of assumptions about the use and characteristics of the vaccine. Even with vaccination, infected and at risk farms would still be likely to be culled¹⁵; we therefore take as our baseline model vaccination of cattle, combined with culling of IPs and epidemiologically identified DCs arising from them⁷. The success of the strategy is defined by the size of the epidemic as a function of the number of farms or animals successfully vaccinated before the outbreak starts (Fig. 1a). The simplest case is vaccination of cattle on a given number of farms chosen at random irrespective of location, size or other risk factors (red line). The resulting number of cases drops relatively rapidly with the number of vaccinated farms; vaccinating about 25,000 cattle farms reduces the size of the epidemic to that achieved by stringent IP, DC and CP culling⁷ (dashed horizontal line). Vaccinating above 80,000 farms prevents almost all major epidemics although, due to the stochasticity of the epidemic process, a precise threshold is not determinable.

Figure 1b considers the same models as Fig. 1a, but shows the expected duration of the epidemic. High vaccine uptake significantly reduces the length of the outbreak by breaking chains of transmission, thus preventing percolation of the infection. This is in contrast with the results of non-spatial deterministic and stochastic

models (see Supplementary Information and Box 1).

These results assume high vaccine efficacy (90%); however, as efficacy may vary considerably, it is important to assess the implications of lower values¹⁴. The inset to Fig. 1a considers vaccine efficacies of 90, 70 and 50%—if we express vaccine uptake in terms of the number of animals protected rather than the number of farms, the resulting epidemic size and duration (not shown) are largely independent of efficacy. This result is similar to standard theoretical predictions^{10,11,16}; essentially, we can compensate for low efficacy by increasing uptake and thereby maintaining the number of animals protected. Because our model epidemics tend to last less

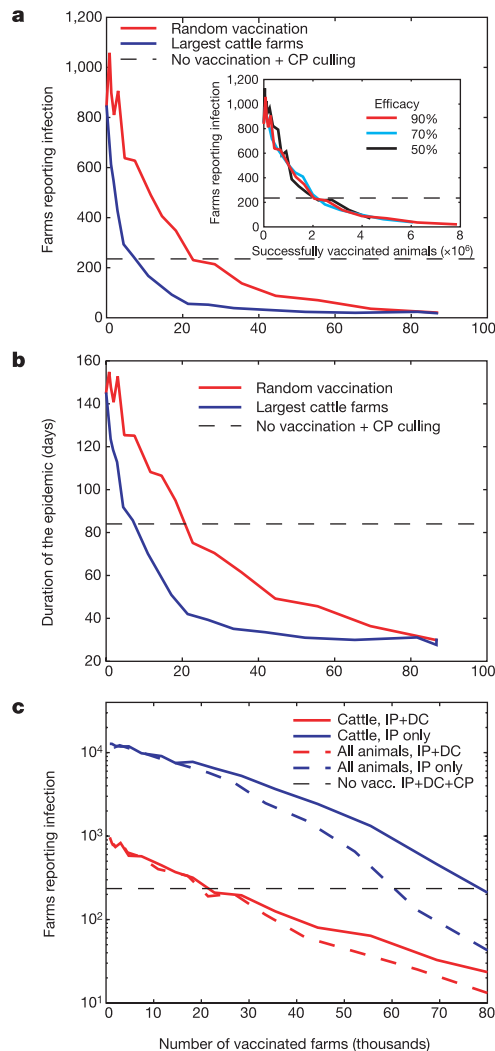


Figure 1 The fate of an introduction of foot-and-mouth disease (FMD) into five randomly chosen farms in Britain after mass vaccination. Infected premises (IP) and dangerous contact (DC) culls (averaging one DC per IP) were used to control the epidemics, and movement restrictions were enforced two weeks after the arrival of the disease. A comparison is made to the case where contiguous premises (CP) culling (70% of CPs are culled) is used instead of vaccination. (All results are the average of 500 simulations, in many of which the epidemic fails to take off.) **a**, Expected number of farms reporting infection against the number of farms vaccinated. Only cattle are vaccinated and the efficacy is assumed to be 90%. The inset figure compares three levels of efficacy for random vaccination. **b**, Expected duration of the epidemic against number of farms vaccinated. The same three vaccination strategies as in **a** are used, vaccinating cattle with 90% efficacy. **c**, Expected number of farms reporting infection when farms are vaccinated at random. Dashed lines indicate that all animals are vaccinated against FMD, whereas solid lines correspond to vaccinating cattle only. Red curves indicate when both IP and DC culls are performed; blue curves are when only IP culls are used.

than 6 months, we do not model explicitly the problem of duration of protection provided by vaccination (Supplementary Information).

Figure 1c explores a range of other prophylactic vaccination options and epidemiological models. First, we consider vaccination of sheep and pigs as well as cattle; this would be more difficult in practice but is useful to explore in terms of the impact of herd immunity. Unsurprisingly, vaccinating all three species (dashed line) produces smaller epidemics than vaccinating cattle only; however, in terms of the total numbers of animals vaccinated, concentrating on cattle alone is much more efficient. Further simulations (see Supplementary Information) indicate that vaccinating sheep only is not an efficient strategy for the FMD strain responsible for the 2001 epidemic in Great Britain, requiring approximately three times as many animals to be vaccinated to achieve the same effect. Sheep-only vaccination also generates a much more variable outcome, as there remains a significant probability of major cattle epidemics (Supplementary Information). Standard theory (Box 1) indicates that targeting of high-risk groups can significantly increase the efficiency of vaccination^{8,10,17}. This is also true here: large cattle farms are particularly susceptible to FMD infection^{6,7} and targeting them (see Methods) significantly decreases the size of the epidemic for a given level of vaccine uptake. In practice, regional targeting of control, informed by risk maps⁷, would probably further improve this performance.

For all-species vaccination, a major epidemic is prevented at around 30% vaccine uptake (40,000 farms). This equates roughly to the results of simple deterministic models⁸, which predict that the effective reproductive ratio, R , of 1.45 after movement restrictions⁴ would require 31% ($=1 - 1/R$) protection to prevent a major epidemic, given that basic FMD control measures are implemented as well. Figure 2 explores this correspondence further by plotting the standard vaccination threshold⁸ against results from the full simulations. The analytical approximation (dashed line) captures the basic shape of the curve calculated for the full system (red) reasonably well; again targeting vaccination at cattle only on large cattle farms (blue) proves more efficacious. A more detailed comparison of epidemics generated by spatial simulations and a simple structured model is given in Box 1. The correspondence between these models is revealing—the deterministic approximation captures the essential shape of the spatial stochastic epidemic, but there is some overestimation of cases during the initial decline of the epidemic. This discrepancy is due to the spatial correlation of high-risk cattle farms, so that vaccination has a larger than anticipated impact in the already infected areas. The tail of the outbreak relies on occasional long-distance jumps to new susceptible areas, and so is underestimated by the non-spatial model; however, the overall shape of the epidemic is partly determined by interspecies heterogeneities in transmission.

Efficient DC culling superimposed on vaccination is important in terms of reducing the epidemic (Fig. 1c, comparison between red and blue lines). This is similar to the general result that prompt local detection and elimination of potential infection is a vital aspect of disease control^{3,4,6,7,18}. Finally, we have so far focused on the epidemiological situation in the 2001 epidemic of Great Britain where pigs were not heavily involved; however, this is frequently not the case^{19,20}. An increased level of pig transmission—reflecting the very high levels of aerosol spread from pigs infected with certain strains of FMD²¹—could result in a much larger epidemic that requires far higher levels of vaccination to control (Supplementary Information). This illustrates the importance of considering a range of infection parameters and the risk of basing strategic planning on the outcome of a single epidemic model.

Reactive vaccination

We consider three approaches to reactive vaccination. First, we assess the feasibility of rapidly achieving herd immunity after

detection of an outbreak—mass reactive vaccination. If this is not possible, we explore more limited strategies of neighbourhood vaccination: ring vaccination, where vaccination is targeted locally in a ring around identified sources of infection²²; and predictive vaccination, where farms that are expected to contribute most to future spatial transmission of infection are targeted.

Of the three approaches to reactive vaccination, we focus on the most likely and most epidemiologically efficient model: mass vaccination of cattle, coupled with IP and DC culling. After the detection of an FMD outbreak, control is governed by two logistical constraints: (1) the ensuing delay before vaccination can be started; and (2) the number of animals that can be vaccinated per day (see Supplementary Information). We assume that sufficient mobilization of vaccine production and veterinary effort could allow vaccination to begin within a week and reach substantial proportions of Great Britain's cattle herd within days or a few weeks after that (see Supplementary Information). We adopt these estimates as the best case and also explore a more pessimistic set of

longer pre-vaccination delays and slower subsequent vaccination rates. Vaccination is applied nationally, but to improve efficiency the largest cattle farms are targeted first (see Methods).

Figure 3a (solid lines) depicts the average size of the simulated epidemic, which declines rapidly with daily vaccination rate, reaching a lower plateau at a rate of around 300,000 cattle per day. This rate corresponds to achieving the deterministic vaccination threshold in around 25 days. The delays, which subsume the time between detection of cases and delivery of vaccine and between vaccination and protection, also have a significant effect on vaccine effectiveness (Fig. 3a); increasing lags allow progressively more of the epidemic to 'escape' the campaign and therefore raise the epidemic size reached even at very high levels of vaccine uptake. Figure 3b (solid lines) shows the corresponding impact of mass reactive vaccination on the duration of the epidemic. Essentially, high levels of herd immunity in cattle can prevent the long tail of the epidemic. Note that relatively short initial delays (less than one month) do not destroy this effect, as later levels of herd immunity

Box 1

Vaccination and control in a heterogeneous environment

Earlier work on vaccination programmes against infectious diseases of humans usually treated populations as homogeneous, and thus correctly estimated high levels of vaccination coverage as being necessary for eradication in most cases. Stimulated largely by work on HIV/AIDS, more recent work has focused on vaccination or other preventive measures in populations with substantial heterogeneity in exposure to, and/or transmission of, infection. One very general conclusion emerges from these studies: if such heterogeneous populations are incorrectly assumed to be homogeneous, then the fraction needing vaccination to eradicate infection will be underestimated (often severely) if we vaccinate at random. By contrast, the eradication fraction can be reduced (often markedly) if we vaccinate optimally, taking advantage of the heterogeneities⁸.

This Box sketches a simple 'toy' model, which gives some insight into the situation where culling and vaccination are combined to control an outbreak of infectious disease among livestock, in such a heterogeneous situation.

Understanding and predicting the dynamics of FMD is complicated by two forms of heterogeneities that have important but conflicting roles. Spatial heterogeneity and the local nature of transmission mean that the number of susceptibles around an infected premise is rapidly depleted. This negative spatial correlation between infected and susceptible farms will reduce the speed with which the epidemic spreads^{3,28}. Farm-level heterogeneities arise due to the differences in number and composition of livestock. Previous analysis has indicated that large farms are both more susceptible and more transmissible than smaller farms, and that cattle farms are more susceptible than sheep farms⁷. This form of heterogeneity can be understood using the modelling framework developed for the heterogeneous transmission of HIV and other sexual diseases⁸.

If each farm has a susceptibility σ and a reproductive potential ρ (transmission rate $\times$ infectious period), then the naive homogenous approach (equivalent to assuming that all farms are identical) would suggest that the basic reproductive ratio and critical vaccination (V_C) level necessary for eradication are:

$$R_0^1 = \langle \rho \rangle \langle \sigma \rangle \approx 1.25 \rightarrow V_C^1 = 1 - \frac{1}{\langle \rho \rangle \langle \sigma \rangle} \approx 20\%$$

Here the value between angle brackets denotes the average value of a quantity across all farms, and the approximate values come from the estimated parameters from the 2001 epidemic in Great Britain and the 2000 farm census data, and ignore spatial heterogeneity. A more complete analysis shows that the basic reproductive ratio is generally larger (as susceptibility and reproductive potential are expected to be

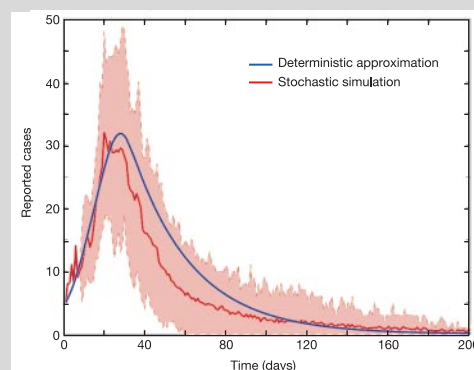
positively correlated), and correspondingly the level of vaccination needed is also increased,

$$R_0^2 = \langle \rho \sigma \rangle = \langle \rho \rangle \langle \sigma \rangle + \text{cov}_{\rho\sigma} \approx 3.60 \rightarrow V_C^2 = 1 - \frac{1}{\langle \rho \sigma \rangle} \approx 72\% > V_C^1$$

where $\text{cov}_{\rho\sigma}$ is the covariance between ρ and σ . However, this higher vaccination threshold assumes that the vaccine is evenly distributed across all farms irrespective of their size. A far lower eradication threshold is achieved if vaccination is targeted; a better but not necessarily optimal policy is to vaccinate farms of type i with probability

$$V_i^3 = \begin{cases} 1 - \frac{1}{\rho_i \sigma_i} & \text{if } \rho_i \sigma_i > 1 \\ 0 & \text{otherwise} \end{cases} \rightarrow \langle V^3 \rangle \approx 18\% < V_C^1$$

This formulation works by reducing the fraction of susceptibles in those groups that are most responsible for transmitting infection. This structured population approach can also be extended to examine mass vaccination in the case where the reduction in susceptible numbers is due primarily to immunization rather than infection or culling. The number of expected cases, I , can be approximated by $I(t+1) \approx I(t) \exp(\langle \tau \sigma \rangle - g)$ where τ is the transmission rate, σ is the susceptibility, g is the slaughter rate of infected livestock, and all three parameters vary to mimic the slaughtering and vaccination policy. The graph below shows the results of 100 stochastic simulations of the full spatial model, with IP culls and mass vaccination (at 300,000 cattle per day), together with the results of the simple model, which has been parameterized to capture the vaccination of largest cattle farms first (see Supplementary Information). These very different models show a considerable degree of correspondence, although this is far from complete.



are key to preventing the tail. Thus, taking into account caveats about FMD strain variation as well as the need for effective vaccines⁷ and the logistics to deliver them (Supplementary Information), prompt mass reactive vaccination of cattle on a large scale could significantly reduce the size and duration of future, disseminated FMD epidemics in Great Britain.

The results discussed above assume effective DC and IP culling in addition to vaccination. The dashed lines in Fig. 3 give the equivalent results assuming no DC culling—this leads to a significantly larger and longer epidemic. There are two components to this effect. First, the initial spread of infection before the start of vaccination is greater if DCs are not culled; second, DC culling is important for reducing R in the unvaccinated sheep population, as well as in unprotected cattle. The latter effect depends on the exact value of R in sheep (and any decisions about sheep vaccination); however, these results underline the importance of effective identification and removal of high-risk premises.

For the neighbourhood vaccination approach, we first consider ring vaccination²² of cattle, again assuming 90% efficacy (see Methods). Figure 4a shows epidemic size as a function of the number of vaccinated farms, which itself varies with the size of the vaccinated ring. Ring vaccination generates some benefit, but even 10-km rings only reduce the epidemic size by around 20%. This lack of effectiveness occurs for the following reasons. First, the delay from infection to reporting and from vaccination to protection means that many local cases arise before vaccination has any effect. Second, the localized nature of ‘IP-centred’ ring vaccination means that neighbouring uninfected areas retain high levels of susceptibility, which can generate new epidemics by means of long-range ‘sparks’ of infection. Ring vaccination is also much less efficacious if combined with CP culling (red), because the latter removes potentially infected and at-risk premises much more rapidly than local vaccination. Further results (see Supplementary Information) indicate that ring vaccination has relatively little effect on the epidemic tail—again, because it leaves susceptible areas where the infection can percolate.

The final approach of reactive vaccination that we consider is predictive vaccination. Limited ring vaccination does not have a strong effect on either the main body or the tail of the epidemic. Only prompt culling or mass vaccination can ‘get ahead’ of the main outbreak; however, if the latter strategies were difficult to implement, a shortening of the epidemic tail (with concomitant economic benefits) could in principle be achieved using a ‘smarter’ predictive vaccination strategy which targets high risk premises.

We demonstrate this potential with the following strategy (see

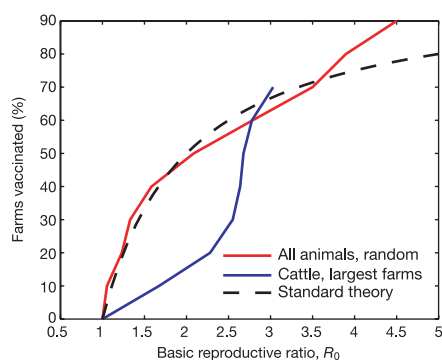


Figure 2 Comparison between simulation (red and blue lines) and standard theory (dashed line) for the critical level of vaccination needed to prevent an epidemic as the basic reproductive ratio, R_0 , varies. Initial conditions are as in Fig. 1. The simulation lines show the level of vaccination needed such that, once movement restrictions are enforced, the average rate of change in the number of cases is zero and hence $R = 1$. Different values of R_0 are modelled by changing all the species-specific transmission rates.

Methods and Box 2). Consider a central ‘source’ farm in which infection has just been detected (Box 2). We use a subsidiary model to predict the probability that a given farm will be infected from this source in the next disease generation and, on the basis of this, the probability of ‘second generation’ infection on each farm. As vaccination takes roughly a disease generation to become effective, first generation farms are not the optimal target as they are already likely to be infected. Instead we should focus control in this predictive horizon on those farms most at risk in the second generation. Beginning vaccination 26 days after the detection of FMD, the predictive strategy gives relatively little benefit in terms of the overall epidemic size compared with ring vaccination (Fig. 4b); however, the duration of the epidemic can be significantly shortened by the predictive strategy, whereas ring vaccination has much less effect (Fig. 4c). This is further demonstrated in Fig. 5a, where expected epidemic curves show that predictive vaccination of just 100 farms per day significantly truncates the epidemic, whereas the addition of CP culling lowers the total epidemic size. Figure 5b shows that if predictive vaccination were applied earlier (8 days after detection) then it also has the potential to significantly lower the deterministic component of the epidemic, whereas even late uptake of vaccination (40 days after detection) can still reduce the duration of the tail. Note that even stringent DC and CP culling cannot achieve this truncation of the outbreak⁷.

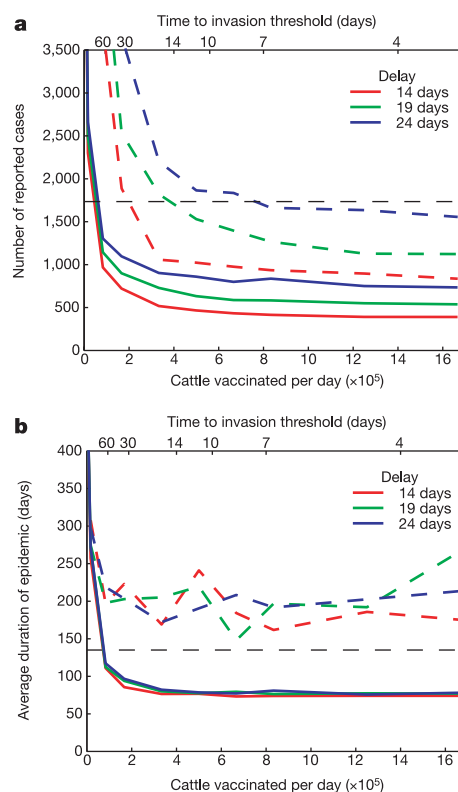


Figure 3 Graph to show how an epidemic can be controlled by the rapid vaccination of cattle during the early stages, using the 2001 epidemic of Great Britain as a template. Throughout, vaccination is of cattle only and assumed to be at 90% efficacy—initially large cattle farms are targeted, as these are most at risk. **a**, Expected number of farms reporting infection against the number of cattle vaccinated per day (bottom axis) or the corresponding time to achieve the disease eradication threshold of around 5.5 million cattle (top axis). Solid lines show the result when IP and DC controls are used; dashed lines show the result when only IP culling is performed. The horizontal line gives the expected number of cases if CP culling is also used but no vaccination is performed. Three different patterns are simulated corresponding to different delays between the detection of the disease and the start of protection by vaccination. **b**, The expected duration of the epidemic for the six patterns investigated in **a**.

Discussion

Our central findings concerning levels of herd immunity and the potential impact of predictive vaccination seem robust to key model assumptions and the limitations of the input data (see Supplementary Information). However, as outlined below (and discussed in the Supplementary Information), these conclusions must confront additional caveats before they can inform policy. Key areas for discussion are: (1) trade issues and cooperation from livestock owners; (2) the practical consequences of vaccination for clinical surveillance, follow-up vaccination to protect newborn animals, vaccination of sheep or pigs in addition to cattle, and the control implications of seasonality in livestock husbandry; and (3) the key issue of understanding the spatial spread of infection (which will vary both with the pathogen strain, its host specificity, farming

practices and the spatial distribution and movement of livestock).

The prophylactic vaccination results mirror a basic epidemiological tenet⁸—raising regional herd immunity sufficiently to reduce the effective reproductive ratio, R , below unity will protect from major epidemics of FMD. Given the observed reproductive ratio of the disease after the imposition of movement restrictions in Great Britain in 2001, immunization of more than 30% of farms would achieve this effect; moreover, targeted vaccination of at-risk groups, such as large cattle farms, could be more effective per immunization. We stress, however, that the prophylactic herd immunity result does not take account of the problem of strain variation in FMD.

Mass reactive vaccination once FMD has been detected and the strain identified could also be effective. Such control would require a rapid and high uptake of vaccination, necessitating cooperation from farmers, and considerable logistical effort and preparation (see Supplementary Information). However it also depends on four other factors. First, a movement ban needs to be instigated promptly, to minimise dissemination of the infection and reduce its reproductive ratio²³. Second, effective detection and removal of livestock on at-risk premises is an important adjunct to vaccination. Third, the birth of large numbers of cattle and sheep at particular times of the year may mean that supplementary vaccination campaigns become necessary. Last, more transmissible strains of FMD virus would require a much greater effort.

The main problem for reactive vaccination is the delay before protection^{14,22}. This particularly affects limited ring vaccination, where the infection spreads to local premises before the vaccination can protect them, allowing ‘escape’ from the edge of the ring. Predictive vaccination is the most efficient limited reactive strategy;

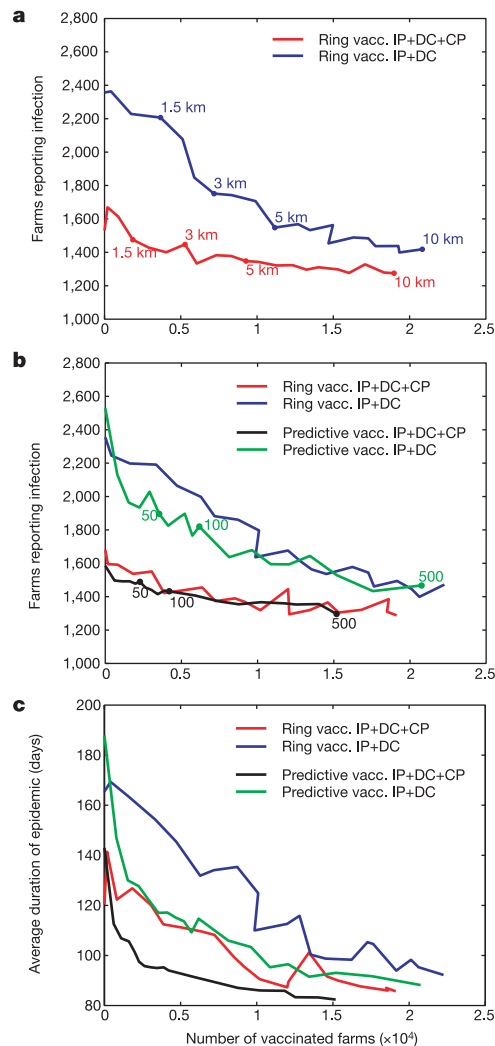


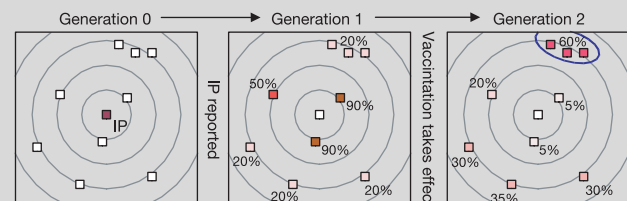
Figure 4 Results from predictive vaccination in comparison to ring vaccination and other control measures. Initial conditions come from the 2001 epidemic in Great Britain, and in general vaccination begins 26 days after disease detection (equivalent to 18 March in the 2001 epidemic). Two different culling policies are considered, either culling DCs and CPs at the observed temporally varying levels⁷, or culling DCs only where the culling level rapidly rises to one DC per IP. **a**, The expected number of farms reporting infection against the total number of farms vaccinated during the epidemic. All non-culled farms within a fixed radius of the IP are vaccinated. Radii of 1.5, 3, 5 and 10 km are indicated. **b**, As for **a**, but with the results for predictive vaccination superimposed. Each day a fixed number of the most at-risk farms are vaccinated; vaccinating levels of 50, 100 and 500 per day are indicated. **c**, Comparison between ring and predictive vaccination in terms of the average duration of the epidemic.

Box 2

Predictive vaccination

Schematic diagrams showing the probability of being infected in each generation. In generation 0 the central farm is infected (it is an IP), whereas the surrounding farms are assumed susceptible. At the end of generation 1, the animals in the IP show signs and the farm reports infection—at this stage vaccination should occur. In generation 1, farms are infected in relation to their distance from the IP. The two farms that are closest have very high probabilities of already being infected, and therefore are unlikely to be infected in the next generation. It is the cluster of farms (circled blue) that have the greatest chance of being infected in generation 2 because: there is a 99% chance that at least one of them is still susceptible in generation 1; they can get infected in generation 2 from farms close to the original IP; and there is a 49% chance that at least one of them got infected in generation 1, subsequent spread to the remainder in generation 2 is then likely. Therefore, in response to the IP, the circled cluster of farms should be vaccinated with the highest priority as here the vaccine has the maximal effect—vaccinating the nearest farms would be futile as they will already be infected before the vaccine takes effect.

The efficacy of predictive vaccination depends on how reliably risk factors for infection and transmission can be identified. Here we have assumed complete knowledge of those risk factors (although the actual outcome is still stochastic in nature). The effectiveness of the strategy will be less if risk factors are less well known.



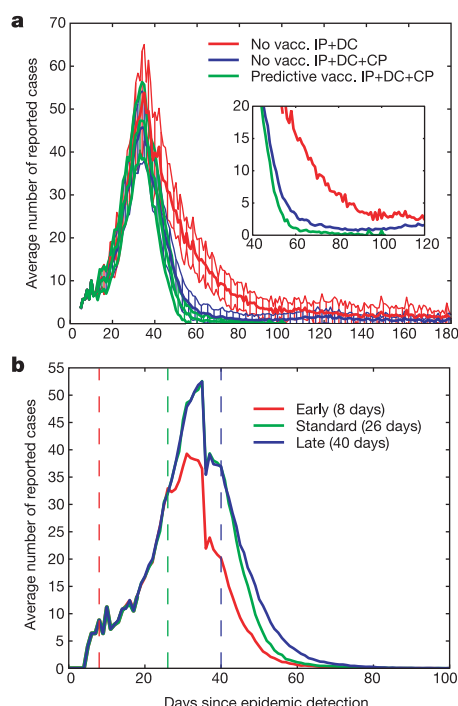


Figure 5 Effects of predictive vaccination on the average epidemic profile, using the same initial conditions as in Fig. 4. **a**, Comparison between culling DCs only, culling DCs and CPs, and culling DCs and CPs in addition to predictive vaccination at a rate of 100 farms per day. Solid lines show the average number of daily cases, with two standard deviations indicated. The inset shows the period around the start of the epidemic tail in greater detail. **b**, The effect of starting predictive vaccination at different dates. The usual start date of 26 days after detection is shown in green, and this is compared to an earlier start of day 8 (red) and a later start of day 40 (blue).

it is essentially an extension of the targeting of at-risk farms with spatial clustering of susceptible farms and the location of infection added as risk factors (Box 2). The method works by blocking transmission to high-risk second-generation nodes in the farm network, potentially reducing the tail of simulated outbreaks by many months. This strategy needs greater epidemiological input before it could be applied: first, to refine our appreciation of risk factors for infection; second, although the method essentially identifies the most at-risk farms, this information then needs to be combined with local expertise to direct the distribution of vaccination.

The optimal strategy depends critically on the initial dissemination of the epidemic. Our analysis is based on the relatively pessimistic (and therefore conservative) model of a highly disseminated epidemic; consequently, we predict that a widespread mass vaccination campaign would be necessary to prevent a major epidemic. Clearly, a single point source epidemic might permit a more limited regional response by ring vaccination; however, our results stress the importance of increasing herd immunity over a relatively wide area to inhibit stochastic jumps of infection for a given spatial transmission kernel.

More broadly, our results probe the robustness of simple mathematical models for eradication of disease by mass regional or national vaccination campaigns⁸. Efficient mass vaccination effectively damps down the spread of infection at all scales; we show that, even in a complex spatial landscape, this effect can be approximated by standard ‘mass action’ models. However, this correspondence only applies to the threshold for eradication of infection by vaccination: lower levels of vaccination can generate complex, nonlinear, spatio-temporal disease dynamics. The pattern here is provided by endemic childhood infections, such as measles and

pertussis⁸, where the interaction between vaccination, epidemic dynamics, seasonal forcing and metapopulation structure can generate a range of subtle dynamical effects^{24–26}.

Turning to local control, this is the first exploration, to our knowledge, of predictive vaccination for optimal control of an infection in a complex spatial landscape. Moreover, our conclusions contradict those obtained using much simpler models for ring vaccination²⁷. The approach will repay further analytical work to optimise control for a given transmission kernel shape and spatial distribution of hosts—a particularly interesting question here is how to optimally reduce the ‘small world’ nature of rare long-range dispersal, which can spark off damaging new local outbreaks during the tail of the epidemic. The spatial focusing of vaccination effort also mirrors, in a more quantitative context, empirical strategies for identifying at-risk individuals in human vaccination campaigns⁸.

There is considerable scope for the development of predictive strategies in the control of emergent human and animal infections. We stress that a prerequisite for success is collaboration between empirical and theoretical researchers and field epidemiologists to have suitable models and relevant databases in place as much as possible before disease outbreaks begin. □

Methods

The basic model

The model used throughout this paper is a spatial stochastic simulation, where the infectious state of every livestock farm in Britain is predicted on a daily basis. The rate, r_i , at which farm i (which is currently susceptible) is infected is given by,

$$R_i = \sum_{L \in \text{livestock}} S_L N_L^i \times \sum_{j \in \text{infectious}} \sum_{L \in \text{livestock}} T_L N_L^j \times K(d_{ij})$$

where N_L^i is the number of livestock of type L within farm i ; S_L is the susceptibility of livestock L ; T_L is the transmission rate of livestock L ; d_{ij} is the distance between farms i and j ; and K is the transmission kernel. Once infected, farms are assumed to remain in an exposed (but not infectious) state for 4 days, after which they become infectious and can transmit the virus to other farms. Nine days after infection, after the appearance of clinical signs, the presence of the disease is reported; after a further delay of between 1 and 3 days (depending of the stage of the epidemic) the animals on the infected farm are slaughtered and the appropriate neighbourhood cull is performed (see Supplementary Information). It was estimated that around 40% of dangerous contacts were infected. More details of the parameter estimation and model validation can be found elsewhere⁷.

Prophylactic vaccination

Prophylactic vaccination is implemented by reducing the number of livestock in those farms vaccinated. In the simplest model, farms are selected at random for vaccination and the outcome is determined independently for each animal. As such the number of susceptible animals remaining in a farm after vaccination is given by a binomial distribution. When vaccinating cattle only, those farms without cattle are ignored. A targeted model is also simulated, where the cattle on those farms with the most cattle are vaccinated first.

Simulations are started with the infection of five randomly chosen farms; often these farms may be quite small and in isolated areas so that the disease dies out. The infection is allowed to spread unchecked for two weeks, after which time movement controls and culling measures are implemented.

In addition to vaccination, all models assume that IPs are culled within 24 h. In most of the simulations culling of DCs is also performed, this is assumed to occur 2 days after reporting, and within one month asymptotes to a ratio of one DC culled per IP. Finally, all simulations are compared to the situation where vaccination is not implemented, but CP culling is; the level of CP culling mimics that achieved during the 2001 epidemic in Great Britain.

Reactive vaccination

Reactive vaccination is implemented using the 2001 epidemic in Great Britain as a template. All simulations start 3 days after the detection of the disease (equivalent to the 23 February in the 2001 epidemic, when movement restrictions were enforced), and use the estimated distribution of infection at this time as initial conditions²³. The timing and level of culls also follows the 2001 pattern. Vaccination is applied to cattle only and is optimistically assumed to offer protection after 4 days; those farms infected before protection is complete were assumed to progress through the normal sequence of infection. Mass vaccination is begun between 1 week and 17 days after the detection of cases. Limited vaccination is generally assumed to start 26 days after detection (equivalent to 18 March in the 2001 epidemic, when it was realized that stronger control measures were needed), and occurs in response to an infected premise 2 days after it is reported.

For the ring vaccination approach, all those farms within a fixed distance of an infected premise will be vaccinated. Note that owing to the nature of the census data this is taken as the distance between farmhouses, rather than the distance between livestock holdings. Vaccinating within the ring was assumed to occur 2 days after the triggering IP is reported, with the entire ring being vaccinated within 24 h (although a further 4 days is needed

before protection is achieved); this is clearly an optimistic assumption and will depend on local logistical constraints.

For the predictive vaccination approach, a fixed number of farms are vaccinated each day, these are chosen to minimise the number of subsequent cases. Using the farms reporting infection so far, we can approximate the infectious state of the system 9 days ago. This is only an approximation, as those farms that are infected but culled before clinical signs arise are not recognized as potential sources of infection. Each week, multiple simulations of a subsidiary model use the reported cases so far to calculate the probability that each susceptible farm becomes infected between 6 and 13 days time. Those farms infected within 6 days would not be protected by vaccination, whereas those farms infected after 13 days could be protected by next week's vaccination. Therefore the prediction procedure requires multiple replicates of a 22-day (9 plus 13) period to determine the infection probability. The farms are then ranked according to the expected decrease in transmission that would be caused by vaccination of their cattle herds.

Received 17 June; accepted 4 December 2002; doi:10.1038/nature01343.

Published online 22 December 2002.

1. Woolhouse, M. & Donaldson, A. Managing foot-and-mouth—The science of controlling disease outbreaks. *Nature* **410**, 515–516 (2001).
2. Gibbens, J. C. *et al.* Descriptive epidemiology of the 2001 foot-and-mouth disease epidemic in Great Britain: the first five months. *Vet. Record* **149**, 729–743 (2001).
3. Ferguson, N. M., Donnelly, C. S. & Anderson, R. M. Transmission intensity and impact of control policies on the foot and mouth epidemic in Great Britain. *Nature* **413**, 542–548 (2001).
4. Woolhouse, M. *et al.* Foot-and-mouth disease under control in the UK. *Nature* **414**, 258 (2001).
5. Morris, R. S., Wilesmith, J. W., Stern, M. W., Sanson, R. L. & Stevenson, M. A. Predictive spatial modelling of alternative control strategies for the foot-and-mouth disease epidemic in Great Britain, 2001. *Vet. Record* **149**, 137–144 (2001).
6. Ferguson, N. M., Donnelly, C. S. & Anderson, R. M. The foot-and-mouth epidemic in Great Britain: pattern of spread and impact of interventions. *Science* **292**, 1155–1160 (2001).
7. Keeling, M. J. *et al.* Dynamics of the 2001 UK foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape. *Science* **294**, 813–817 (2001).
8. Anderson, R. M. & May, R. M. *Infectious Diseases of Humans* (Oxford Science, Oxford, 1991).
9. Pastoret, P. P. *et al.* (eds) *Veterinary Vaccinology* (Elsevier, Amsterdam, 1997).
10. McLean, A. R. & Blower, S. M. Modelling HIV vaccination. *Trends Microbiol.* **3**, 458–463 (1995).
11. Woolhouse, M. E. J., Haydon, D. T. & Bundy, D. A. P. The design of veterinary vaccination programmes. *Vet. J.* **153**, 41–47 (1997).
12. Williams, J. R., Nokes, D. J. & Anderson, R. M. Targeted hepatitis B vaccination—a cost effective immunisation strategy for the UK? *J. Epidemiol. Comm. Health* **50**, 667–673 (1996).
13. Woolhouse, M. E. J. *et al.* Failure of vaccination to prevent outbreaks of foot-and-mouth disease. *Epi. Infect.* **116**, 363–371 (1996).
14. Barnett, P. V. & Carabin, H. A review of emergency foot-and-mouth disease (FMD) vaccines. *Vaccine* **20**, 1505–1514 (2002).
15. EU strategy for emergency vaccination against foot-and-mouth disease (FMD). Report of the Scientific Committee on Animal Health and Animal Welfare (<http://europa.eu.int/comm/dg24/>

health/sc/scah/index_en.html) (1999).

16. Woolhouse, M. E. J. & Bundy, D. A. P. in *Veterinary Vaccinology* (eds Pastoret, P. P. *et al.*) 565–573 (Elsevier, Amsterdam, 1997).
17. Woolhouse, M. E. J. *et al.* Heterogeneities in the transmission of infectious agents: implications for the design of control programs. *Proc. Natl Acad. Sci. USA* **94**, 338–342 (1997).
18. HMSO Northumberland Report. *The Report of the Committee of Inquiry on Foot and Mouth Disease 1968* (HMSO, London, 1968).
19. Howard, S. C. & Donnelly, C. A. The importance of immediate destruction in epidemics of foot and mouth disease. *Res. Vet. Sci.* **69**, 189–196 (2000).
20. Yang, P. C. *et al.* Epidemiological characteristics and financial costs of the 1997 foot-and-mouth disease epidemic in Taiwan. *Vet. Record* **145**, 731–734 (1999).
21. Donaldson, A. I. *et al.* Relative risks of the uncontrollable (airborne) spread of FMD by different species. *Vet. Record* **148**, 602–604 (2001).
22. Salt, J. S. *et al.* in *Veterinary Vaccinology* (eds Pastoret, P. P. *et al.*) 641–652 (Elsevier, Amsterdam, 1997).
23. Haydon, D. T. *et al.* The construction and analysis of epidemic trees with reference to the 2001 UK foot-and-mouth outbreak. *Proc. R. Soc. Lond. B* (in the press); published online (doi:10.1098/rspb.2002.2191).
24. Keeling, M. J., Rohani, P. & Grenfell, B. T. Seasonally-forced disease dynamics explored as switching between attractors. *Physica D* **148**, 317–335 (2001).
25. Rohani, P., Earn, D. J. D. & Grenfell, B. T. Opposite patterns of synchrony in sympatric disease metapopulations. *Science* **286**, 968–971 (1999).
26. Grenfell, B. T., Bjørnstad, O. N. & Kappey, J. Travelling waves and spatial hierarchies in measles epidemics. *Nature* **414**, 716–723 (2001).
27. Muller, J., Schonfisch, B. & Kirkilionis, M. Ring vaccination. *Math. Biol.* **41**, 143–171 (2000).
28. Keeling, M. J. The effects of local spatial structure on epidemiological invasions. *Proc. R. Soc. Lond. B* **266**, 859–869 (1999).

Supplementary Information accompanies the paper on Nature's website

(<http://www.nature.com/nature>).

Acknowledgements This research was supported by the Royal Society (M.J.K.) and the Wellcome Trust (M.E.J.W. and B.T.G.). Data were provided by the Department of the Environment, Food and Rural Affairs, the Scottish Executive Rural Affairs Department, and the National Assembly for Wales Statistical Directorate. We thank D. Shaw, S. St Rose, M. Chase-Topping, D. Haydon and L. Matthews for their help.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.J.K. (e-mail: m.j.keeling@warwick.ac.uk).

An early lunar core dynamo driven by thermochemical mantle convection

Dave R. Stegman*, A. Mark Jellinek*†, Stephen A. Zatman‡§, John R. Baumgardner|| & Mark A. Richards*

* Department of Earth and Planetary Science, University of California, Berkeley, California 94720, USA

† Department of Physics, University of Toronto, Toronto, Canada M5S 1A7

‡ Department of Earth and Planetary Sciences, Washington University, St Louis, Missouri 63130, USA

|| Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

§ Deceased

Although the Moon currently has no internally generated magnetic field, palaeomagnetic data, combined with radiometric ages of Apollo samples, provide evidence for such a magnetic field from ~3.9 to 3.6 billion years (Gyr) ago¹, possibly owing to an ancient lunar dynamo^{1,2}. But the presence of a lunar dynamo during this time period is difficult to explain^{1–4}, because thermal evolution models for the Moon⁵ yield insufficient core heat flux to power a dynamo after ~4.2 Gyr ago. Here we show that a transient increase in core heat flux after an overturn of an initially stratified lunar mantle might explain the existence and timing of an early lunar dynamo. Using a three-dimensional spherical convection model⁶, we show that a dense layer, enriched in radioactive elements (a 'thermal blanket'), at the base of the lunar mantle can initially prevent core cooling, thereby inhibiting core convection and magnetic field generation. Subsequent radioactive heating progressively increases the buoyancy of the thermal blanket, ultimately causing it to rise back into the mantle. The removal of the thermal blanket, proposed to explain the eruption of thorium- and titanium-rich lunar mare basalts⁷, plausibly results in a core heat flux sufficient to power a short-lived lunar dynamo.

Table 1 Model parameters

Symbol	Description	Value
R_o	Radius of planet	1,740 km
R_i	Radius of core	450 km
L	Thickness of blanket layer	230 km
H	Depth of mantle	1,290 km
ρ_m	Density, mantle	3,400 kg m ⁻³
ρ_c	Density, core	7,400 kg m ⁻³
g_s	Gravity, surface of Moon	1.6 m s ⁻²
g_c	Gravity, surface of lunar core	0.93 m s ⁻²
α_m	Thermal expansivity, mantle	4 × 10 ⁻⁵ K ⁻¹
α_c	Thermal expansivity, core	10 ⁻⁴ K ⁻¹
k_m	Thermal conductivity, mantle	4 W m ⁻¹ K ⁻¹
k_c	Thermal conductivity, core ^{30,31}	25–50 W m ⁻¹ K ⁻¹
$C_{p,m}$	Heat capacity, mantle	1,400 J kg ⁻¹ K ⁻¹
$C_{p,c}$	Heat capacity, core	850 J kg ⁻¹ K ⁻¹
L_m	Latent heat, mantle	5 × 10 ⁵ J kg ⁻¹
T_o	Surface temperature	250 K
$T_c(t=0)$	Initial core temperature (models TB-0,1,2)	1,800 K
$T_s(0)$	Solidus temperature at surface, mantle	1,450 K
$T_s(P)$	Solidus temperature increase, mantle	0.1 K Mpa ⁻¹
η_o	Reference viscosity, mantle	3 × 10 ²⁰ Pa s
E	Activation energy	5 × 10 ⁵ J mol ⁻¹
V	Activation volume	1.5 × 10 ⁻⁶ m ³ mol ⁻¹
a	Scaling for viscosity exponent	0.28
γ	Maximum viscosity variation	3,000
[U]	Uranium concentration	25.7 p.p.b.
Th/U	Thorium/uranium ratio	4
K/U	Potassium/uranium ratio	2,000
f	Fraction of radiogenic elements in mantle	0.25

$H(t)$ is heat production rate for bulk lunar mantle due to radioactive element concentrations listed above and their appropriate decay constants. Assuming that radiogenic elements are concentrated into the last 5% of mantle volume during magma ocean solidification, the heat production of the KREEP layer is $20H(t)$. During overturn, some fraction f of the KREEP layer is entrained into this downwelling flow and the resultant thermal blanket layer is assumed to have a uniform heat-production of $20fH(t)$.

Solidification of an early planetary-scale lunar magma ocean (see, for example, ref. 8) produced a chemically stratified mantle. An anorthosite crust was underlain by the last-solidified 'KREEP' (enriched in potassium, rare-earth elements and phosphorus) layer, containing most of the mantle's incompatible, radioactive heat-producing elements. Beneath the KREEP layer was an earlier-formed ilmenite cumulate⁷ overlying a thick layer of pyroxenite, which formed the bulk of the lunar mantle. The proposed ilmenite cumulate layer is denser than the underlying pyroxenite⁷. Thus, shortly after solidification of the lunar mantle, the ilmenite cumulate layer overturned and sank into the mantle as a number of drips, ponding as a layer at the core–mantle boundary⁹ (CMB). Some fraction of the KREEP layer must also have been viscously entrained into these downwellings. This dense, mixed ilmenite cumulate layer overlying the CMB would have contained a high concentration of radioactive elements. Subsequent radioactive decay and internal heating of this layer might have caused it to become thermally buoyant and rise back into the overlying mantle^{7,10}. Without the downward transport of heat-producing elements provided by an initial overturn, a purely conductive model could reasonably describe the lunar thermal evolution¹¹.

We propose that the removal of this dense, internally heated layer (thermal blanket) overlying the core caused a transient increase in the core heat flux that was sufficient to power a lunar dynamo. However, to explain the observed palaeomagnetic intensity data, three criteria must be satisfied. First, the core cooling must be sufficiently rapid to drive thermal convection in the core^{3,12–15}. Second, convective motions in the core must supply enough magnetic energy to power a strong-field dynamo^{12–15}, the regime in which terrestrial planetary magnetic fields are thought to be generated¹⁶. Third, this dynamo must produce magnetic field strengths at the lunar surface that are compatible with the palaeointensity data^{1–3} (Fig. 1i).

Thermal convection inside a molten iron core occurs when the core heat flux exceeds that conducted down the core adiabat^{12–15}, $Q_{ad} = \alpha_c T_c k_c g_c / c_{p,c}$ (see Table 1 for symbol definitions and values). Although both the initial temperature and composition of the core are poorly constrained, the current presence of a fluid core¹⁷ implies that the core is not pure iron but instead contains a low-density impurity such as sulphur, which has lowered its freezing temperature. The solidification of a pure-Fe inner core from a liquid Fe–S alloy probably began early in the evolution of the lunar core and drove compositional convection¹³. Gravitational potential energy supplied by compositional convection would increase, by as much

Table 2 Model runs

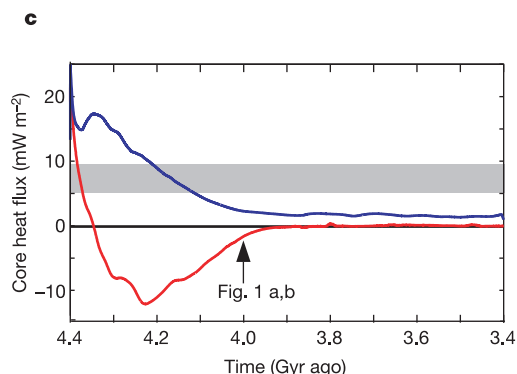
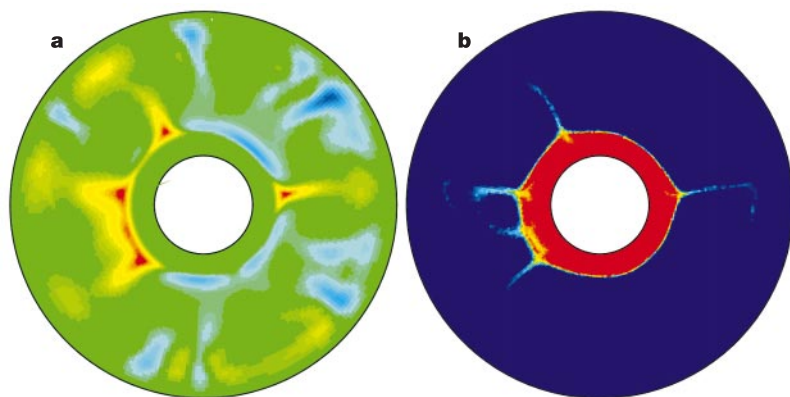
Case	Buoyancy number, B	Initial temperature drop across CMB, ΔT (K)
TB-0	0	200
TB-1	1.0	200
TB-2	0.5	200
TB-3	0.5	0

We use a massively parallel, finite element model (LUNA), adapted from TERRA^{6,18,19}, with a lagrangian particle tracer method³² to accommodate the heterogeneous distribution of chemically dense material, a smaller core relative to the planet's radius (the temperature of which can evolve according to the conservation of heat at the core–mantle boundary) and time-dependent, heterogeneous heat production. The model also accounts for the conservation of latent heat on melting throughout the depth of the mantle. The computational mesh resolution is 10 km at the inner surface and 30 km at the outer surface and each element initially contains 10 particles, which carry chemical-density and heat-production information. The core radius is 450 km, which for technical reasons associated with the numerical grid employed is the smallest size currently possible but is just outside the range of observational constraints on lunar core size. Because mantle convection occurs beneath a 'stagnant lid'^{20,21}, we impose a no-slip boundary condition at the surface and a free-slip condition at the core–mantle boundary. The radial viscosity of lunar mantle depends on temperature according to $\eta(T) = \eta_o e^{(E+PV)/RT}$, where T is the laterally averaged temperature at each depth and a is a damping constant introduced for technical reasons that limits the exponential temperature dependence of viscosity¹⁹. Even with the more modest sensitivity to temperature in these models, there is still a viscosity variation in the calculations, γ (measured as the ratio of largest to smallest viscosities), equal to 3,000, most of which is in the upper thermal boundary layer. Ignoring lateral viscosity variations affects the detailed form of upwellings and downwellings but still allows us to capture the essential thermal evolution of the models, as well as the basic nature of the hot and cold thermal boundary layers³³. The ΔT used in B is the temperature drop across the interface between ambient mantle and the thermal blanket layer, as determined in case TB-1 (ref. 24). The chosen B is only an initial value, because B evolves during the calculation.

as an order of magnitude, estimates for the energy available to drive a dynamo based on thermal convection alone^{13,14}. However, this additional energy depends on the surface area of the inner core. Because both the sulphur concentration and initial temperature in

the core are poorly constrained it is difficult to estimate the size of an early inner core. Thus, we are unable to assess the role of compositional convection. Our initial goal is, therefore, to identify conditions in which thermal evolution models predict thermal

Stable stratification: model TB-1



Unstable stratification: model TB-2

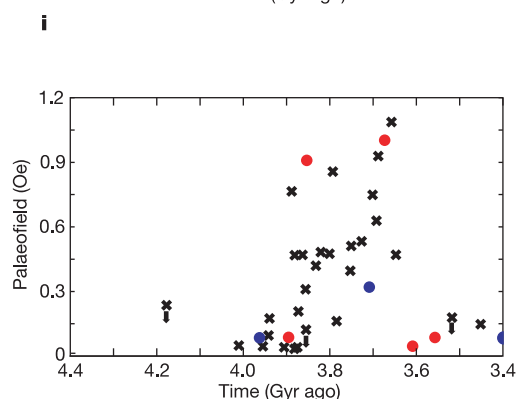
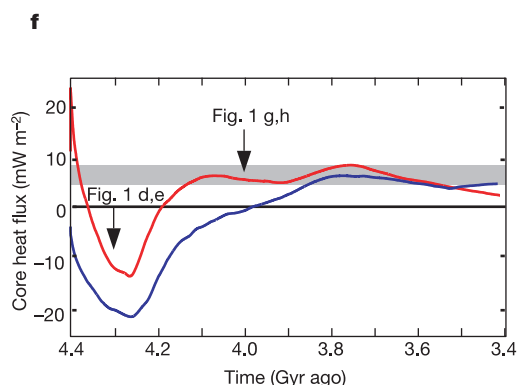
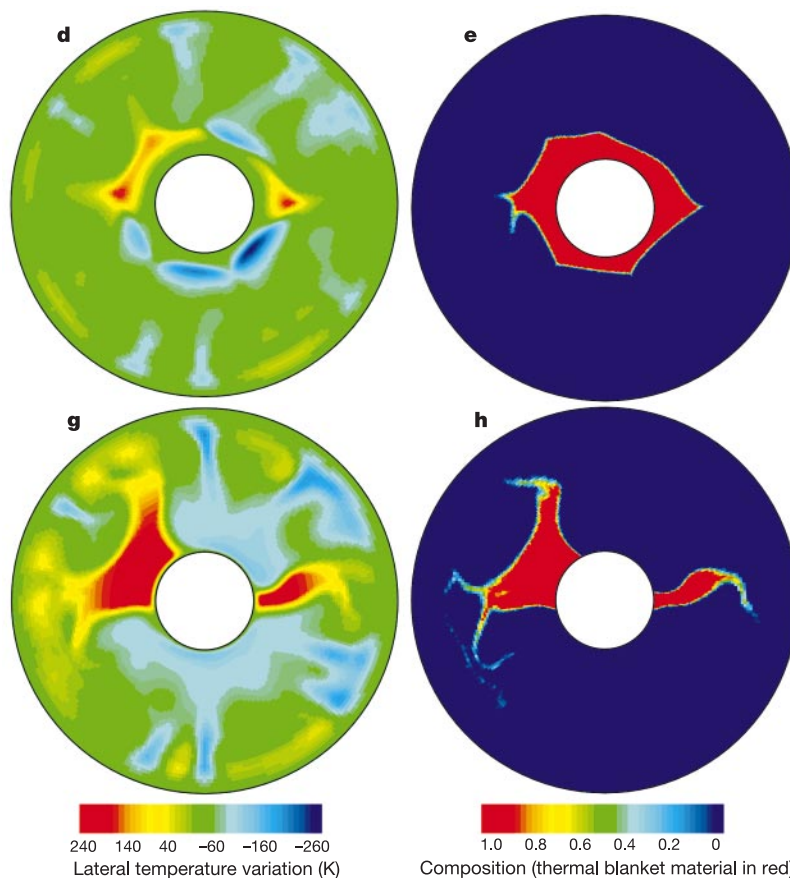


Figure 1 Thermochemical evolution models in stable (a–c) and unstable (d–h) thermal blanket regimes as seen in temperature (a, d, g), composition (b, e, h) and core heat flux (c, f) compared with palaeomagnetic intensity data (i). The equatorial cross-sections of temperature (a) and composition (b) for model TB-1 at 4.0 Gyr ago show that thermal blanket material is too dense to become buoyant, but some entrainment occurs. However, equatorial cross-sections of temperature and composition for model TB-2 show a marginally stable thermal blanket interacting with mantle convection at 4.3 Gyr ago (d, e) and that by 4.0 Gyr ago (g, h) has sufficient thermal buoyancy to rise back towards the lunar surface. c, The core thermal history for reference model TB-0 shows heat flux values (blue line) well below adiabatic core heat flux (grey region), whereas model TB-1 (red line) has nearly zero heat flux (black line). Such core heat flux values ranging between the black line and grey region indicates a thermally stratified core in which all core heat loss is by

conduction and no dynamo is supported. f, Thermal blanket ($B = 0.5$). Red line, initial ΔT at CMB = 200 K; blue line, initial ΔT at CMB = 0 K. A core heat flux equal to or above the shaded region indicates core convection and the probable occurrence of a dynamo, as seen in models TB-2 and TB-3. i, Palaeointensity measurements from Apollo samples (modified from ref. 1) in which dots indicate absolute palaeointensity measurements (Thellier–Thellier method in red; other techniques in blue) and crosses indicate scaled normalized relative palaeointensities. In our models, asymmetric removal of the thermal blanket leads to a localized distribution of partly molten thermal-blanket material at relatively shallow mantle depths, confirming a plausible explanation for the eruption of high-thorium, high-titanium mare basalts, similar to the models in ref. 10. Our models make no attempt to evaluate melt transport to the surface. $1 \text{ Oe} \approx 79.6 \text{ A m}^{-2}$.

convection in the core (that is, $Q_{\text{cmb}} > Q_{\text{ad}}$).

Internal radioactive heating and cooling at the Moon's surface result in mantle convection. To model early lunar mantle convection, we use a three-dimensional spherical numerical model described in detail elsewhere^{6,18,19}. Reasonable lunar mantle properties (Tables 1, 2) yield a Rayleigh number, $Ra = \rho g \alpha \Delta T H^3 / \mu \kappa = 3 \times 10^6$, appropriate for the early Moon. Because the lunar crust is thought to have remained intact from shortly after the Moon's formation, mantle convection has evidently occurred in the 'stagnant-lid' regime, in which large viscosity variations due to the temperature dependence of mantle viscosity results in an immobile upper thermal boundary layer, or lithosphere, and a thin sublayer subject to convective instability^{20,21}. This sublayer has a viscosity that is about an order of magnitude larger than the underlying mantle²¹, corresponding to a temperature difference of ~ 100 K. Such weak cooling, characteristic of stagnant-lid convection, leads to unfavourable conditions for dynamo generation²².

Our models begin ~ 100 Myr after Moon formation, during which time solidification of the magma ocean occurs, dense material ponds at the base of the mantle, and solid-state mantle convection begins. Initially the dense layer is 230 km thick and has a uniform density indicated by the buoyancy number, $B = \Delta \rho_{\text{chem}} / \alpha_m \Delta T$. The temperature profile beneath the upper thermal boundary layer is that of the mantle solidus to 400 km depth. Below this depth, the temperature is taken to be a constant 1,600 K, similar to that in other studies^{5,10}. We examine initial core temperatures 0–200 K hotter than the overlying mantle⁵ (Table 2).

Model TB-0 is a reference thermal evolution case⁵ in which the core cools to a mantle of uniform density and composition (no thermal blanket). Core temperatures range from 1,600 to 1,700 K, giving $Q_{\text{ad}} = 5.0\text{--}9.0 \text{ mW m}^{-2}$. Typical of one-plate planets, the core heat flux, Q (blue line in Fig. 1c), exceeds Q_{ad} (shaded region) only for the first few hundred million years, an effect of the core's beginning 200 K hotter than the mantle. Such simple thermal evolution models⁵ fail to generate core convection after about ~ 4.2 Gyr ago.

In model TB-1 (Fig. 1a–c), $B = 1.0$ and the thermal blanket is stable^{23,24}. The thermal blanket heats the outermost portion of the core (red line in Fig. 1c). At about 4 Gyr ago, several hot plumes rise from the top of the thermal blanket into the overlying mantle entraining small amounts of dense material. After ~ 4 Gyr ago, the core cools more slowly than for TB-0. $Q < Q_{\text{ad}}$ except in the very earliest stages of evolution (~ 4.4 Gyr ago).

In model TB-2, $B = 0.5$ and the thermal blanket is marginally stable^{23,24}. Initially, as in model TB-1, the thermal blanket warms the lunar core (Fig. 1d and e). However, prolonged internal radioactive heating increases the buoyancy of the thermal blanket, so that it ultimately rises into overlying mantle as a few large upwelling plumes¹⁰ (Fig. 1g and h). The removal of the thermal blanket causes $Q \geq Q_{\text{ad}}$ from ~ 4.0 to 3.5 Gyr ago, coincident with the magnetic era (~ 3.9 to 3.6 Gyr ago). Even in the case for which the core and lowermost mantle begin in thermal equilibrium, model TB-3 (blue line in Fig. 1f), evolves such that $Q \geq Q_{\text{ad}}$.

Our results show that all simple layered (TB-1) or unlayered (TB-0) models lead to core heat fluxes an order of magnitude too small to drive thermal convection in the core during the magnetic era. However, in the marginally stable regime^{23,24} (TB-2), the removal of a thermal blanket owing to internal heating and dynamic coupling with mantle convection results in a core heat flux sufficient to drive thermal convection in the core for a few hundred million years beginning ~ 4 Gyr ago. Furthermore, a key prediction of model TB-2 is that a well-defined onset time for the lunar dynamo, approximately coincident with thermal blanket removal, might be observable $\sim 4.2\text{--}3.9$ Gyr ago (Fig. 1f).

Whether model TB-2 results in a strong-field dynamo¹⁶ depends on the efficiency with which the available gravitational potential energy driving thermal convection is converted into magnetic

energy through ohmic dissipation (see ref. 14, for example). For a core of radius of 350–450 km (refs 17, 25), ohmic dissipation must exceed $\sim 30\text{--}70$ MW to support the strong-field regime^{14,16}. Applying the lunar core properties in Table 1 to the study of Buffett *et al.*¹⁴ indicates that, during the period that model TB-2 has a super-adiabatic core heat flux, the ohmic heat dissipation is $\sim 0.1\text{--}0.2\%$ of the total core heat flow. This efficiency implies that $\sim 20\text{--}40$ MW is available for ohmic dissipation and thus a strong-field dynamo is plausible. Additional forcing by compositional convection^{13,14} would further increase the energy available to drive a dynamo.

The surface magnetic field strength resulting from model TB-2 is difficult to constrain. The magnetic field is partitioned into two components: a toroidal component that remains in the core and a poloidal component observable at the surface. The partitioning of magnetic field energy between these two components is unknown. However, poloidal magnetic field strength is proportional to $1/R^3$. If model TB-2 supports a 12-G (1 gauss = 10^4 tesla) core field (strong-field dynamo) by thermal convection alone, and the poloidal component represents about half of that, the maximum surface field strength would be ~ 0.1 G. To achieve even stronger surface fields (0.2–0.5 G), consistent with those observed during the magnetic era (Fig. 1i), core field strengths of 20–60 G are needed. This suggests compositional convection driven by the growth of an inner core was an important process during the magnetic era.

The results of model TB-2 might also explain the origin of the recently recognized Procellarum KREEP Terrane (PKT)^{11,26,27}, a province defined by its elevated levels of thorium, which has undergone extensive resurfacing by mare volcanism. Petrological analyses of volcanic glasses²⁸ support the inferences from remotely sensed data that high-thorium, high-titanium basalts exist within the PKT^{27,29}, although Apollo basalt samples from this region are low in thorium¹¹. Moreover, the depths of origin of these glasses can exceed 500 km (refs 27–29) and might share a mantle source with PKT basalts. If the high-titanium source region originally solidified at a shallow depth (< 100 km)^{7,8}, a cumulate overturn model^{7,10} might account for such deep melting nearly 500 Myr later. Zhong *et al.*¹⁰ reported that although the proposed overturn was a global event, the ilmenite cumulate material that had ponded at the CMB probably rose back to the surface asymmetrically, producing localized high-thorium, high-titanium volcanism. Our results are consistent with thermal blanket removal's being an inherently asymmetric process and thus provide a self-consistent explanation for both the existence of an ancient lunar dynamo as well as the origin, composition and duration of mare volcanism in the PKT. □

Received 25 June; accepted 25 October 2002; doi:10.1038/nature01267.

1. Cisowski, S. M., Collinson, D. W., Runcorn, S. K. & Stephenson, A. A review of lunar paleointensity data and implications for the origin of lunar magnetism. *J. Geophys. Res.* **88**, A691–A704 (1983).
2. Collinson, D. W. Magnetism of the Moon—A lunar core dynamo or impact magnetization? *Surv. Geophys.* **14**, 89–118 (1993).
3. Stevenson, D. J. Planetary magnetic fields. *Rep. Prog. Phys.* **46**, 555–620 (1983).
4. Runcorn, S. K. The formation of the lunar core. *Geochim. Cosmochim. Acta* **60**, 1205–1208 (1996).
5. Konrad, W. & Spohn, T. Thermal history of the Moon: Implications for an early core dynamo and post-accretional magmatism. *Adv. Space Res.* **19**, 1511–1521 (1997).
6. Baumgardner, J. R. Three dimensional treatment of convective flow in the Earth's mantle. *J. Stat. Phys.* **39**, 501–511 (1985).
7. Hess, P. C. & Parmentier, E. M. A model for the thermal and chemical evolution of the Moon's interior: Implications for the onset of mare volcanism. *Earth Planet. Sci. Lett.* **134**, 501–514 (1995).
8. Warren, P. H. The magma ocean concept and lunar evolution. *Annu. Rev. Earth Planet. Sci.* **13**, 201–240 (1985).
9. Jellinek, A. M., Kerr, R. C. & Griffiths, R. W. Mixing and compositional stratification produced by natural convection. 1. Experiments and their application to Earth's core and mantle. *J. Geophys. Res.* **104**, 7183–7202 (1999).
10. Zhong, S., Parmentier, E. M. & Zuber, M. T. A dynamic origin for the global asymmetry of lunar mare basalts. *Earth Planet. Sci. Lett.* **177**, 131–140 (2000).
11. Wiczeorek, M. A. & Phillips, R. J. The 'Procellarum KREEP Terrane': Implications for mare volcanism and lunar evolution. *J. Geophys. Res.* **105**, 20417–20430 (2000).
12. Stevenson, D. J., Spohn, T. & Schubert, G. Magnetism and thermal evolution of the terrestrial planets. *Icarus* **54**, 466–489 (1983).
13. Stacey, F. D. & Loper, D. E. Thermal histories of the core and mantle. *Phys. Earth Planet. Inter.* **36**, 99–115 (1984).

14. Buffett, B. A., Huppert, H. E., Lister, J. R. & Woods, A. W. On the thermal evolution of the Earth's core. *J. Geophys. Res.* **101**, 7989–8006 (1996).
15. Stevenson, D. J. Mars' core and magnetism. *Nature* **412**, 214–219 (2001).
16. Roberts, P. H. & Soward, A. M. Dynamo theory. *Annu. Rev. Fluid Mech.* **24**, 459–512 (1992).
17. Williams, J. G., Boggs, D. H., Yoder, C. F. & Ratcliff, J. T. Lunar rotational dissipation in solid body and molten core. *J. Geophys. Res.* **106**, 27933–27968 (2001).
18. Bunge, H.-P. & Baumgardner, J. R. Mantle convection modeling on parallel virtual machines. *Comput. Phys.* **9**, 207–215 (1995).
19. Yang, W.-S. & Baumgardner, J. R. A matrix-dependent transfer multigrid method for strongly variable viscosity infinite Prandtl number thermal convection. *Geophys. Astrophys. Fluid* **92**, 151–195 (2000).
20. Christensen, U. R. Heat transport by variable viscosity convection and implications for the Earth's thermal evolution. *Phys. Earth Planet. Inter.* **35**, 264–282 (1984).
21. Moresi, L.-N. & Solomatov, V. S. Numerical investigation of 2D convection with extremely large viscosity variations. *Phys. Fluids* **7**, 2154–2162 (1995).
22. Jellinek, A. M., Lenardic, A. & Manga, M. The influence of interior mantle temperature on the structure of plumes: Heads for Venus, tails for Earth. *Geophys. Res. Lett.* **29**, doi:10.1029/2001GL014624 (2002).
23. Davaille, A. Simultaneous generation of hotspots and superswells by convection in a heterogeneous planetary mantle. *Nature* **402**, 756–760 (1999).
24. Gonnermann, H. M., Manga, M. & Jellinek, A. M. Dynamics and longevity of an initially stratified mantle. *Geophys. Res. Lett.* **29**, doi:10.1029/2002GL01485 (2002).
25. Hood, L. L., Mitchell, D. L., Lin, R. P., Acuna, M. H. & Binder, A. B. Initial measurements of the lunar induced magnetic dipole moment using Lunar Prospector magnetometer data. *Geophys. Res. Lett.* **26**, 2327–2330 (1999).
26. Jolliff, B. L., Gillis, J. J., Haskin, L. A., Korotev, R. L. & Wiczorek, M. A. Major lunar crustal terranes: Surface expressions and crust–mantle origins. *J. Geophys. Res.* **105**, 4197–4216 (2000).
27. Haskin, L. A., Gillis, J. J., Korotev, R. L. & Jolliff, B. L. The materials of the lunar Procellarum KREEP Terrane: A synthesis of data from geomorphological mapping, remote sensing, and sample analyses. *J. Geophys. Res.* **105**, 20403–20415 (2000).
28. Delano, J. Pristine lunar glasses: Criteria, data and implications. *J. Geophys. Res.* **91**, 201–213 (1986).
29. Hess, P. C. On the source regions for mare picrite glasses. *J. Geophys. Res.* **105**, 4347–4360 (2000).
30. Anderson, O. L. The Gruneisen parameter for iron at outer core conditions and the resulting conductive heat and power in the core. *Phys. Earth Planet. Inter.* **109**, 179–197 (1998).
31. Stacey, F. D. & Anderson, O. L. Electrical and thermal conductivities of Fe–Ni–Si alloy under core conditions. *Phys. Earth Planet. Inter.* **124**, 153–162 (2001).
32. Stegman, D. R., Richards, M. A. & Baumgardner, J. R. Effects of depth dependent viscosity and plate motions on maintaining a relatively uniform MORB reservoir in whole mantle flow. *J. Geophys. Res.* **107**, doi:10.1029/2001JB000192 (2002).
33. Tackley, P. J. Effects of strongly variable viscosity on 3-dimensional mantle convection. *J. Geophys. Res.* **101**, 3311–3332 (1996).

Acknowledgements We thank B. Buffett for helpful discussions; C. Johnson, R. Jeanloz, M. Manga and H.-P. Bunge for comments; and D. Stevenson and M. Zuber for reviews that improved this manuscript. We dedicate this work to the memory of our co-author Stephen Zetman, who inspired us by his interest in this subject. This work was supported by IGPP LANL, NASA CT project, the National Science Foundation, and the Miller Institute for Basic Research in Science.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.S. (e-mail: dstegman@eps.berkeley.edu).

Measuring intense rotation and dissipation in turbulent flows

Benjamin W. Zeff*, Daniel D. Lanterman*, Ryan McAllister*, Rajarshi Roy*, Eric J. Kostelich† & Daniel P. Lathrop*

* Department of Physics, IPST and IREAP, University of Maryland, College Park, Maryland 20742-4111, USA

† Department of Mathematics, Arizona State University, Tempe, Arizona 85287-1804, USA

Turbulent flows are highly intermittent—for example, they exhibit intense bursts of vorticity and strain. Kolmogorov theory^{1,2} describes such behaviour in the form of energy cascades from large to small spatial and temporal scales, where energy is dissipated as heat. But the causes of high intermittency in turbulence, which show non-gaussian statistics^{3–5}, are not well understood. Such intermittency can be important, for example, for enhancing the mixing of chemicals^{6,7}, by producing sharp

drops in local pressure that can induce cavitation (damaging mechanical components and biological organisms)⁸, and by causing intense vortices in atmospheric flows. Here we present observations of the three components of velocity and all nine velocity gradients within a small volume, which allow us to determine simultaneously the dissipation (a measure of strain) and enstrophy (a measure of rotational energy) of a turbulent flow. Combining the statistics of all measurements and the evolution of individual bursts, we find that a typical sequence for intense events begins with rapid strain growth, followed by rising vorticity and a final sudden decline in stretching. We suggest two mechanisms which can produce these characteristics, depending whether they are due to the advection of coherent structures through our observed volume or caused locally.

The problem of turbulence has been notoriously difficult to resolve, because it involves the interaction of flow length and timescales over many orders of magnitude. The intense, localized events that are characteristic of turbulence can cause difficulty in numerical simulations, via the generation of small scales and large local gradients. Simulations have some advantage in that they provide access to the entire field; alternatively, experiments can yield long observational records that are computationally impractical. Turbulence experiments also have limitations. Many measurements have relied on invasive techniques⁹ that can affect the flow being studied. Usually, measurements do not capture the full three-dimensional nature of the flow; instead, they rely on one- or two-component approximations of velocities and velocity gradients. A fully three-dimensional (3D), spatially resolved technique is needed to measure important quantities such as the energy dissipation and helicity within the flow. Good temporal resolution is also required to study the rapid development of intense events.

A few recent experiments have captured many of these features. Tao *et al.*^{10,11} used holographic particle image velocimetry (HPIV) to take elegant 3D snapshots of turbulent channel flow, though the technique is not time-resolved. Using two very different technologies, La Porta *et al.*¹² and Mordant *et al.*¹³ tracked individual particles (a lagrangian measurement) in highly turbulent flows. This

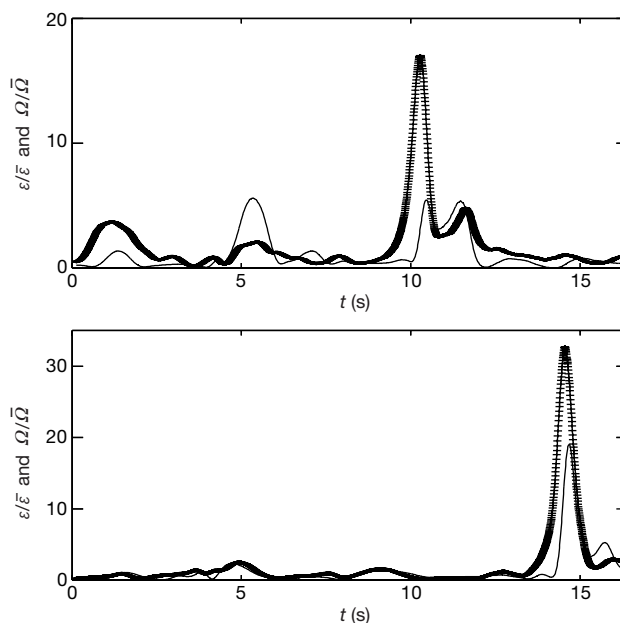


Figure 1 Time traces of the dissipation ε and enstrophy Ω . The two traces above show intense events occurring in a turbulent flow. Dissipation curves are marked with crosses, and enstrophy curves are shown as continuous curves. The time averages are $\bar{\varepsilon} = 0.0336 \text{ cm}^2 \text{ s}^{-3}$ and $\bar{\Omega} = 4.13 \text{ s}^{-2}$ with intermittent bursts reaching up to 40 times those values.

method provides direct access to the velocity and acceleration of the flow, but cannot yield velocity gradients.

We have designed and implemented a 3D particle image velocimetry technique¹⁴ that obtains time series of the three components of velocity and all nine velocity gradients measured within a small volume. That is, this instrument allows us not only to study the statistics of intense, intermittent events, but also to observe and characterize their dynamics. In the following, we describe this technique and the results from a turbulent flow.

To visualize the 3D velocity field locally, we align three laser sheets such that each sheet illuminates one face of a small, cubic test volume, with the three sheets crossing at one vertex of the cube. The sheets illuminate 1 μm spherical polystyrene fluorescent particles that track the motion of the fluid. Three high-speed video cameras, recording at 125 frames per second, focus on the sheets, such that each camera sees only those particles in its own sheet. The key to reconstructing the 3D flow from these three slices is that the test volume be sufficiently small so that the flow within is, to good approximation, locally linear:

$$\mathbf{v} \approx \mathbf{b} + M\mathbf{x} \quad (1)$$

where $M = \partial v_i / \partial x_j$ is the matrix of gradients. We use the two-dimensional velocity vectors from all three faces to estimate $\mathbf{b}$ and M for the whole cube. The dissipation ε and enstrophy Ω can be calculated at a given time step from the measured gradients: $\varepsilon = \nu \Sigma_{ij} (M_{ij} + M_{ji})^2 / 2$ and $\Omega = \Sigma_{ij} (M_{ij} - M_{ji})^2 / 4 = \omega^2 / 2$, where $\omega = \nabla \times \mathbf{v}$ is the vorticity and ν is the kinematic viscosity. In a turbulent flow, for equation (1) to hold, the test region must be of the order of, or smaller than, the Kolmogorov length η , the length scale at which viscosity smooths out the flow. We use three long-distance microscopes to image a small region (of size 1.8η), from a working distance of 43.3 cm .

Our flow is produced in a transparent cubic container of side $l = 22\text{ cm}$. Two square grids with a mesh size of $l_g = 0.85\text{ cm}$ are rigidly connected to each other with a vertical spacing of 6.0 cm and are vertically oscillated with a position $y = y_0 \sin(\omega_g t)$ to drive the flow. The measurement volume at the centre of the tank is

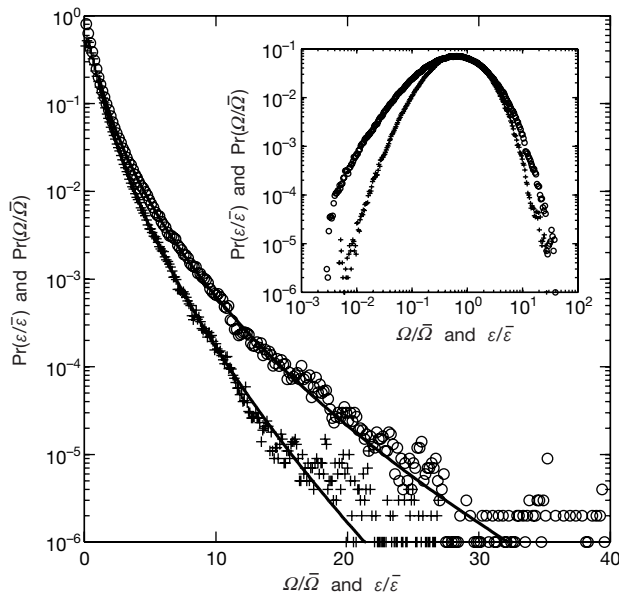


Figure 2 Distributions of the dissipation and enstrophy. The long tails of the probability distribution functions extend toward high values of ε (crosses) and Ω (circles), indicating that these are highly intermittent quantities. For values larger than their means, the distributions are well approximated by stretched exponentials, $\text{Pr}(\varepsilon) \approx \exp[\varepsilon^{-0.62 \pm 0.08}]$ and $\text{Pr}(\Omega) \approx \exp[\Omega^{-0.45 \pm 0.07}]$, as indicated by the solid lines. Inset, log-log plots show significant deviations from log-normality, which would yield parabolas here.

horizontally and vertically centred between the two grids.

The Reynolds number, which gives the approximate ratio of inertial to viscous forces in the flow, can be defined in terms of the tank size and the velocity amplitude of the grid, $\text{Re} = l\omega_g y_0 / \nu$; we present data for a flow with $\text{Re} = 48,200$ and $R_\lambda = v_{x,\text{rms}}^2 / \nu \langle (\partial_x v_x)^2 \rangle^{1/2} = 54$. The time average of the dissipation is $\bar{\varepsilon} = 0.0336\text{ cm}^2 \text{ s}^{-3}$, which can be used to estimate the Kolmogorov length scale for this flow, $\eta = (\nu^3 / \bar{\varepsilon})^{1/4} = 0.074\text{ cm}$.

To gather sufficient statistics, 5×10^6 sets of vectors and matrices for grid turbulence have been gathered, corresponding to 0.8 terabytes of video data. Video sequences of the flow reveal periods of unidirectional flow occasionally interrupted by visually striking vortices and hyperbolic flow patterns. Traces of ε and Ω (Fig. 1) reflect the intermittent nature of the flow. These distributions are qualitatively comparable to those observed in numerical simulations¹⁵. The probability distributions $\text{Pr}(\varepsilon)$ and $\text{Pr}(\Omega)$ (Fig. 2) both display long tails with high values more common for Ω than ε , indicating that vorticity is more intermittent than strain. Each component of the gradient matrix M assumes positive and negative values and, by continuity, must occasionally pass through zero. A one-dimensional approximation $\varepsilon \approx 15\nu(\partial_x u_x)^2$ of these quantities, as used in some previous experiments³, implies that very low values of ε occur with much higher probability than we observe. In reality, it is unlikely that all nine gradient components simultaneously drop to zero.

Taken over the entire data set, the measure $\langle \varepsilon \Omega \rangle = 1.699 \bar{\varepsilon} \bar{\Omega}$ indicates that these two quantities are not statistically independent. A scatter plot of ε versus Ω (Fig. 3) highlights the interdependence of these quantities. In particular, especially high values of ε coincide with high values of Ω sixteen times more frequently than if they were independent.

Our observations of the fluctuating gradients permit a statistical analysis of the interdependencies of the stretching and rotating parts of the field. It is useful to split the gradient matrix M into symmetric (representing strain) and antisymmetric (representing rotation) parts: $M = S + A$. Note that the dissipation is proportional to the sum of the squared components of S and that the enstrophy is proportional to the sum of the squared components of A . We observe that the time derivatives $\dot{\varepsilon}$ and $\dot{\Omega}$ are systematically affected by the local conditions of ε (and the three eigenvalues of S , which we denote as λ_i) and Ω . In particular, the dissipation on average is rising when the largest eigenvalue of S , λ_1 , is large (see Fig. 4a), and falling when the enstrophy is large (see Fig. 4b).

The time dependence of the dissipation can be due to several

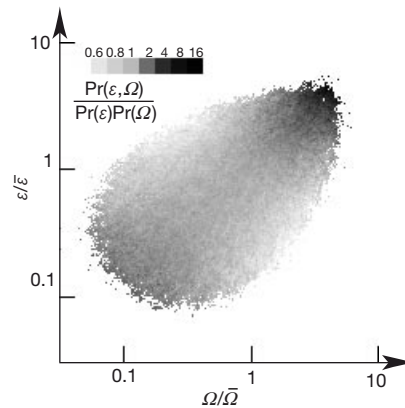


Figure 3 A scatter plot of dissipation versus enstrophy. The interdependence of these two quantities is clearly evident. The shading $\text{Pr}(\varepsilon, \Omega) / \text{Pr}(\varepsilon)\text{Pr}(\Omega)$ would be 1 if ε and Ω are independent. While values close to their averages appear uncorrelated, high values of ε and Ω occur simultaneously about sixteen times more often than would be expected. Particularly low values appear similarly correlated. Only points with $\text{Pr}(\varepsilon, \Omega) > 10^{-5}$ were plotted.

mechanisms. Although we will also advance the idea of local kinematic causes, another explanation invokes coherent structures being advected into our measurement volume. Because our observations are probably an admixture of the two phenomena, we will discuss interpretations relating to both.

Previous observations of vortex tubes with nearby sheets of high dissipation¹⁶ offer one possible explanation to the behaviour seen in Fig. 1. An advecting velocity correlated with these structures is consistent with the observed statistics in the following way. Figure 4b shows the dissipation falling as the enstrophy increases. This is consistent with a dissipation decrease as the centre of a vortex core is approached, and that the high dissipation sheet precedes but does

not trail the vortex. Figure 4c shows that the front of this preceding dissipation sheet would typically be compressive in two directions, while the back would have one compressive direction.

We can learn about gradient fluctuations by studying our data and by examining the Navier–Stokes equations. If we expand the velocity field locally as a first-order Taylor series in $\mathbf{x}$, exactly as in equation (1), and substitute it into the Navier–Stokes equations, we obtain the following equation for M : $\dot{M} = -M^2 - \Pi/\rho$, where Π is the Hessian matrix for the pressure field¹⁷. This expansion neglects several factors, all quadratic in $\mathbf{x}$, including the effect of viscosity and the effect of advection in convecting gradients into our local region. This expression retains gradient steepening by the nonlinear term (M^2) and the nonlocal effect of pressure to enforce incompressibility. Note that Π is undetermined except by boundary conditions and incompressibility, $\text{Tr}(M^2 + \Pi) = 0$; this is due to the non-locality of the pressure. Finally, splitting M into symmetric and antisymmetric parts yields:

$$\dot{S} = -S^2 - A^2 - \Pi/\rho \quad (2)$$

$$\dot{A} = -SA - AS \quad (3)$$

These form a local basis for the interaction of rotation (A) and strain (S). Because the matrix A contains only the components of the vorticity, equation (3) gives $\dot{\omega} = S\omega$, indicating how the strains stretch and amplify vorticity¹⁷.

The kinematics of equations (2) and (3) can be related to our observed statistics and by observations of events such as those displayed in Fig. 1.

First, the gradients show self-steepening instabilities, causing $\dot{\varepsilon} > 0$. These instabilities are probably of the same type that causes the gradients to grow in the Burger's equation. The conditional statistic shown in Fig. 4a substantiates this effect. Second, the presence of large gradients causes the vorticity to grow ($\dot{\Omega} > 0$) owing to the well-known vortex stretching mechanism. Finally, once the vorticity has grown, it causes the dissipation to fall, that is, $\Omega \gg \bar{\Omega}$ implies that $\dot{\varepsilon} > 0$, as is seen in Fig. 4b and anticipated in equation (2) (although the effect of the $-A^2$ term depends in a nontrivial way on the directions involved).

The events shown in Fig. 1 were chosen for their particularly large amplitudes of ε , but they show a sequence bearing out these three features. The gradients (as indicated by ε) first grow, in the absence of a large vorticity. Once the gradients are large, the vorticity (through Ω) shows a sudden, rapid increase. The peaking of the enstrophy is then accompanied by the rapid decline of the dissipation. We see no clear evidence, through conditional statistics, for a mechanism by which the vortices decline. It may be that slower viscous forces are required to damp the vortices, because kinematic mechanisms, as represented in equations (2) and (3), are ineffectual at damping vorticity.

Using this optical technique, we present these experiments to directly observe the development, interdependence and interaction of turbulent strain and vorticity. We have directly quantified intermittent events in the dissipation and enstrophy and discussed their relation to advecting structures. Our analysis includes a characterization of self-steepening and the first statistical evidence for vortex 'squelch' (see Fig. 4) of the dissipation. Our observations highlight how an understanding of the temporal interactions between stretching and vorticity is crucial to the science of extreme events in turbulence. □

Methods

Each image frame was broken into four 108×128 pixel subsections (9.7% overlap). A two-dimensional cross-correlation is performed between each subsection in subsequent images, a process which identifies the best two-dimensional pixel shift to particle images from one frame to the next. A weighted averaging technique is used to identify the cross-correlation peak to sub-pixel resolution, estimated at 0.25 pixels. A linear least-squares method fits the calculated velocity vectors and three-vector locations from each camera to the linear shear model of equation (1) to estimate $\mathbf{b}$ and M . Vector validation and

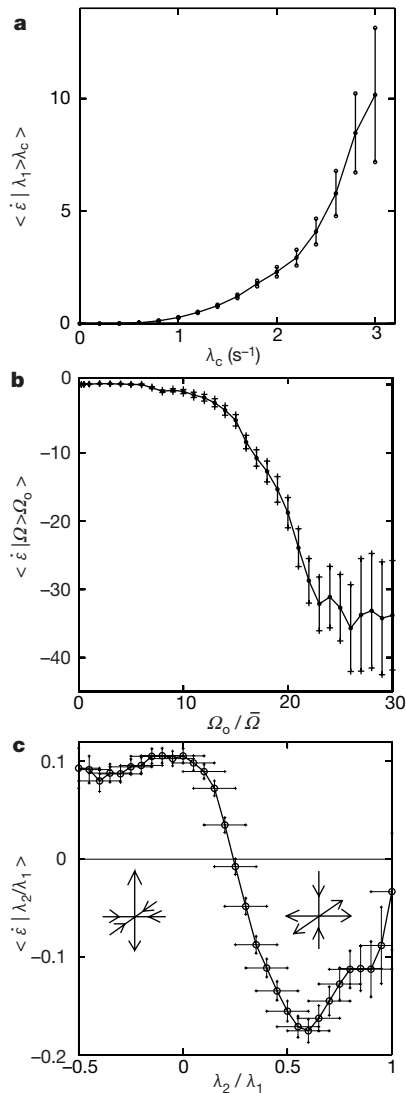


Figure 4 Conditional averages of the growth rate of dissipation $\dot{\varepsilon}$. **a**, The average of $\dot{\varepsilon}$ conditioned on the largest eigenvalue of the strain matrix λ_1 being larger than a value λ_c . Gradients, represented by λ_1 , are subject to a self-steepening mechanism, causing the dissipation to rise. Note that $\varepsilon = 2\nu\Sigma_i\lambda_i^2$. **b**, The average of $\dot{\varepsilon}$ conditioned on the enstrophy Ω being larger than a value Ω_0 . The observations show an unambiguous decline in the dissipation when rotations are strong. We interpret this as a vortex 'squelch', whereby the vortices disrupt by mixing the self-steepening of the straining motions. **c**, The average of $\dot{\varepsilon}$ conditional on the ratio of the two largest eigenvalues of the strain matrix. Arrows indicate local flow stretching directions. The dissipation, on average, grows when there are two contracting directions, while for $\lambda_2/\lambda_1 > 0.25$ the dissipation, on average, falls. As the contracting directions are the ones unstable in growth by a Burger's mechanism, this may indicate that having two contracting directions enhances that mechanism.

bootstrapping¹⁸ are used to reduce the effect of spurious vectors caused by too few particles. The least-squares calculation is constrained to be divergence-free, that is, $\text{Tr}(\mathbf{M}) = 0$.

Received 11 July; accepted 26 November 2002; doi:10.1038/nature01334.

1. Kolmogorov, A. N. The local structure of turbulence in the incompressible viscous fluid for very large Reynolds numbers. *Dokl. Akad. Nauk. SSSR* **30**, 301–305 (1941); reprinted in *Proc. R. Soc. Lond. A* **434**, 9–13 (1991).
2. Kolmogorov, A. N. Dissipation of energy in the locally isotropic turbulence. *Dokl. Akad. Nauk. SSSR* **31**, 538–540 (1941); reprinted in *Proc. R. Soc. Lond. A* **141**, 15–17 (1991).
3. Sreenivasan, K. R. Fractals and multifractals in fluid turbulence. *Annu. Rev. Fluid. Mech.* **23**, 539–600 (1991).
4. Meneveau, C. & Sreenivasan, K. R. Simple multifractal cascade model for fully developed turbulence. *Phys. Rev. Lett.* **59**, 1424–1427 (1987).
5. Frisch, U. *Turbulence: The Legacy of A.N. Kolmogorov* 120–194 (Cambridge Univ. Press, Cambridge, 1995).
6. Gollub, J. P., Clarke, J., Gharib, M., Lane, B. & Mesquita, O. N. Fluctuations and transport in a stirred fluid with a mean gradient. *Phys. Rev. Lett.* **67**, 3507–3510 (1991).
7. Chaté, H., Villersmaux, E. & Chormaz, J.-M. (eds) *Mixing: Chaos and Turbulence* (Kluwer Academic/Plenum, New York, 1999).
8. Kunns, K. T. & Papoutsakis, E. T. Damage mechanisms of suspended animal cells in agitated bioreactors with and without bubble entrainment. *Biotech. Bioeng.* **36**, 476–483 (1990).
9. Vukoslavcic, P., Wallace, J. M. & Balint, J.-L. The velocity and vorticity vector fields of a turbulent boundary layer. Part 1. Simultaneous measurement by hot-wire anemometry. *J. Fluid Mech.* **228**, 25–51 (1991).
10. Tao, B., Katz, J. & Meneveau, C. Geometry and scale relationships in high Reynolds number turbulence determined from three-dimensional holographic velocimetry. *Phys. Fluids* **12**, 941–944 (2000).
11. Tao, B., Katz, J. & Meneveau, C. Statistical geometry of subgrid-scale stresses determined from holographic particle image velocimetry measurements. *J. Fluid Mech.* **457**, 35–78 (2002).
12. La Porta, A., Voth, G. A., Crawford, A. M., Alexander, J. & Bodenschatz, E. Fluid particle accelerations in fully developed turbulence. *Nature* **409**, 1017–1019 (2001).
13. Mordant, N., Metz, P., Michel, O. & Pinton, J.-F. Measurement of lagrangian velocity in fully developed turbulence. *Phys. Rev. Lett.* **87**, 214501 (2001).
14. Keane, R. D. & Adrian, R. J. Theory of cross-correlation analysis of PIV images. *Appl. Sci. Res.* **49**, 191–215 (1992).
15. Chen, S., Sreenivasan, K. R. & Nelkin, M. Inertial range scalings of dissipation and enstrophy in isotropic turbulence. *Phys. Rev. Lett.* **79**, 1253–1256 (1997).
16. Kida, S. & Ohkitani, K. Spatiotemporal intermittency and instability of a forced turbulence. *Phys. Fluids A* **4**, 1018–1027 (1992).
17. Majda, A. J. & Bertozzi, A. L. *Vorticity and Incompressible Flow* 6–13 (Cambridge Univ. Press, Cambridge, 2002).
18. Kostelich, E. J. Bootstrap estimates of chaotic dynamics. *Phys. Rev. E* **64**, 016213 (2001).

Acknowledgements We gratefully acknowledge the support of the National Science Foundation and the Research Corporation. R.M. and R.R. gratefully acknowledge support from the Office of Naval Research (Physics Division). We thank D. Levermore, J. Rodgers, D. DeShazer, K. R. Sreenivasan, E. Ott, T. Antonsen and J. Fineberg for advice.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.P.L. (e-mail: dpl@complex.umd.edu).

Observation of surface and bulk phase transitions in nematic liquid crystals

M. I. Boamfa, M. W. Kim, J. C. Maan & Th. Rasing

NSRIM Institute and High Field Magnet Laboratory, University of Nijmegen, Toernooiveld 1, 6525 ED, Nijmegen, The Netherlands

The behaviour of liquid crystal (LC) molecules near a surface is of both fundamental and technological interest: it gives rise to various surface phase-transition and wetting phenomena^{1–14}, and surface-induced ordering of the LC molecules is integral to the operation of LC displays^{15,16}. Here we report observation of a pure isotropic–nematic (IN) surface phase transition—clearly separated from the bulk IN transition—in a nematic LC on a substrate. Differences in phase behaviour between surface and bulk are expected^{1–4}, but have hitherto proved difficult to

distinguish, owing in part to the close proximity of their transition temperatures. We have overcome these difficulties by using a mixture of nematic LCs: small, surface-induced composition variations lead to complete separation of the surface and bulk transitions, which we then study independently as a function of substrate and applied magnetic field. We find the surface IN transition to be of first order on surfaces with a weak anchoring energy and continuous on surfaces with a strong anchoring. We show that the presence of high magnetic fields does not change the surface IN transition temperature, whereas the bulk IN transition temperature increases with field. We attribute this to the interaction energy between the surface and bulk phases, which is tuned by magnetic-field-induced order in the surface-wetting layer.

In a confined geometry, interactions between LC molecules and the surfaces of the confining walls give rise to different LC behaviour at the surface and in the bulk. Upon undergoing the IN transition the delicate balance of intermolecular interactions tends to favour the LC–surface interaction in the nematic phase, whereas the LC–LC interaction tends to dominate in the isotropic phase. The new nematic phase therefore occurs first at the surface, a phenomenon known as “wetting”^{1,2,13,14}. The IN transition is relatively weak thermodynamically, so large pre-transitional effects take place before the transition. The interaction of these pre-ordered LC molecules with the surface is known as “pre-wetting”^{2,13,14}. In a magnetic field the diamagnetic LC molecules undergo a torque that forces them to orient relative to the field direction¹⁷. Although at the level of individual molecules the magnetic field effect is negligibly small in the isotropic phase, the collective behaviour of the meso-phase causes the molecular contribution to add up constructively, resulting in a macroscopic orientation effect (Fig. 1).

A variety of theoretical studies have addressed this wetting problem either at the general level of surface-related phenomena^{1,2} or for LCs in particular^{3,4}. Cahn¹ introduced a first-order wetting transition leading to an ordering film of finite thickness in any two-phase mixture of fluids contacting a third phase at a critical temperature. Later, Nakanishi and Fisher² showed that pre-wetting of a highly ordered film with finite thickness on a surface continues to wetting via a first-order or continuous wetting transition. This pre-wetting, where a strongly anchored LC layer forms a uniform nematic phase adjacent to the substrate and near the isotropic bulk, has often been discussed for the case of strong surface fields^{3,4}.

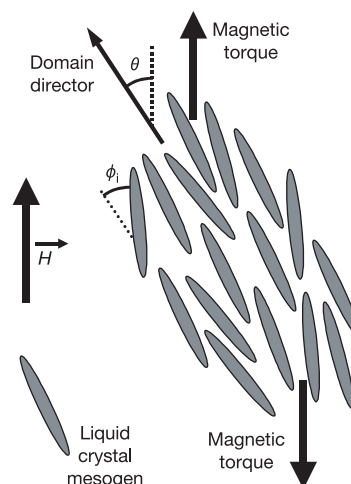


Figure 1 A correlated volume of liquid crystal mesogens under the influence of an applied magnetic field, H . The field tends to orient the domain as a whole, changing the director orientation relative to the field, θ , leaving unchanged the individual molecules within the domain, ϕ_i .

Sluckin and Poniewierski⁴ calculated a general phase diagram of nematic LC in the presence of both bulk LC and a magnetic field. It shows a wetting transition changing from first order to continuous with either increasing surface interaction or increasing magnetic field.

Experimental data to verify this theoretical work are limited. Polarized optical microscopy observations^{5,6} revealed different kinds of behaviour near the phase transition depending on the surface, and ellipsometry^{7–10} and nuclear magnetic resonance¹¹ studies demonstrated surface pre-wetting in the isotropic phase. But until now, no observations have confirmed the existence of a first-order, combined first-order/continuous or continuous surface transition. Experimental difficulties related to the investigation of such transitions include the very close proximity of the surface transition to the bulk transition in all cases studied so far.

As samples we used planar cells, prepared with the polymer-coated (or clean) glass substrates and adjusted to 5.6 μm in thickness, using polyethyleneterephthalate (PET) spacers, filled with the LC mixture E63 (Merck; $T_{\text{IN}} = 87.5^\circ\text{C}$, composed¹⁸ of six LC mesogenes, all diamagnetic) in the isotropic phase. The thermoplastic polymer films used as cell-surface coatings, polyvinyl cinnamate (PVCN) and polyvinyl formal (PVFM), were prepared on a glass substrate from 2% polymer solution by spin coating and subsequent thermal annealing for solvent removal. For experiments the cells were placed in a temperature-controlled chamber inside a 20 tesla (T) Bitter-magnet, with the glass plates parallel to the magnetic field. The field-induced birefringence and the optical transmission of the samples were measured simultaneously with the help of a polarization-intensity double modulated laser beam, He-Ne 543.5 nm, which passed through the sample at normal incidence relative to the magnetic-field direction. We probed the

LC order *in situ* with optical birefringence¹⁷, as well as the LC morphology with light scattering¹⁹. Scattering is the only contribution to optical transmission because the E63 does not absorb light at the wavelength we used.

Figure 2 shows the surface and the bulk IN phase transition on different substrates as the cells were cooled in a magnetic field of 2 T. The surface IN transition appears as a small step in the birefringence baseline ($\Delta b_{\text{IN}}^{\text{s}}$), taking place over a finite temperature range ($\Delta T_{\text{IN}}^{\text{s}}$), a few kelvin above the bulk transition. The non-zero birefringence baseline slope represents the pre-transitional LC ordering. The surface transition is directly related to the molecular interaction between the LC and the substrate. On a clean glass substrate, characterized by a very weak surface anchoring ($G \approx 10^{-10} \text{ J m}^{-2}$), we find that the surface transition has a sharp first-order character and appears at the lowest temperature of all the studied cells (Fig. 2a). For the PVFM substrate, with a medium anchoring energy ($G \approx 10^{-7} \text{ J m}^{-2}$) originating mainly from van der Waals interactions between polar groups, the surface transition occurs at a higher temperature and shows a combined first-order/continuous appearance (Fig. 2b). The PVCN substrate has a strong anchoring ($G \approx 10^{-5} \text{ J m}^{-2}$), because the interaction also includes the π - π coupling with the cinnamoyl groups; the transition therefore occurs at the highest temperature and has a clear continuous character (Fig. 2c). From $\Delta b_{\text{IN}}^{\text{s}}$ (using the full cell birefringence at room temperature as reference), we estimate the thickness of the ordering LC film to be about one molecular layer, confirming the initial prediction of ref. 1 and the observations of ref. 11. Note that this corresponds to a wetting transition on a pre-wetted surface and therefore does not contradict other wetting layer thickness estimations^{13,14}. Figure 3 shows that with increasing magnetic field, the bulk IN transition temperature (T_{IN}^{b}) shifts to higher values, but the surface IN transition temperature (T_{IN}^{s}) as well as its temperature window ($\Delta T_{\text{IN}}^{\text{s}}$) stay unchanged. However, the slope of the birefringence versus temperature dependence, related to the pre-transitional order, gets steeper and the jump in the birefringence corresponding to the surface transition becomes larger for higher applied magnetic fields. This indicates a higher degree of induced alignment, both for the pre-ordered bulk molecules and surface layer, for larger fields. For sufficiently large magnetic fields, T_{IN}^{b} approaches T_{IN}^{s} to the extent that the two transitions are no longer seen as distinct (Fig. 3d).

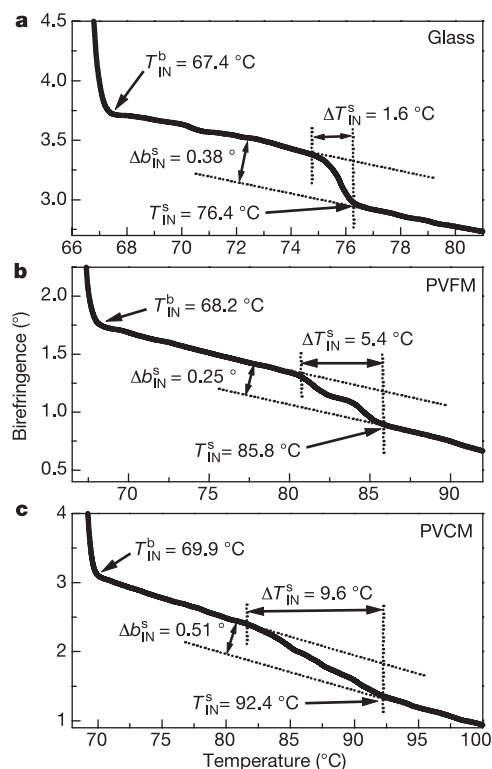


Figure 2 Surface and bulk isotropic–nematic phase transitions on different substrates. **a–c**, Magnetic-field-induced birefringence of E63 on clean glass (**a**), polyvinyl formal (PVFM) (**b**) and polyvinyl cinnamate (PVCN) (**c**) at 2 T. Detail is shown around the surface T_{IN}^{s} and bulk T_{IN}^{b} transition temperatures. $\Delta T_{\text{IN}}^{\text{s}}$ and $\Delta b_{\text{IN}}^{\text{s}}$ are respectively the temperature interval and the field-induced birefringence associated with the surface transition.

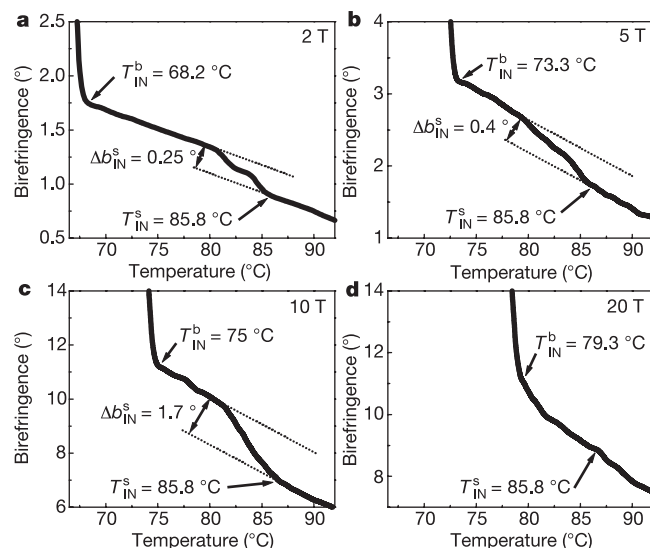


Figure 3 Magnetic-field dependence of the surface and bulk transitions. **a–d**, Field-induced birefringence evolution of a PVFM-coated cell in different magnetic fields: 2 T (**a**), 5 T (**b**), 10 T (**c**) and 20 T (**d**). Detail is shown around the surface and bulk IN transitions; data were taken upon cooling.

Figure 4 presents the birefringence and optical transmission of a PVFM-coated sample that was cooled in a field of 5 T and subsequently warmed at 0 T. For the cooling-down process the birefringence signal shows at first a small increase, corresponding to the pre-translational order, followed by a small step at T_{IN}^{s} (for the detail around T_{IN}^{s} see Fig. 3b). A few degrees below the surface transition step, the birefringence signal shows a sharp increase, indicative of the bulk IN transition. The curve follows the same initial slope until room temperature is reached. Recorded simultaneously, the transmission curve shows a small, continuous decrease through the surface transition interval. This suggests that the surface layer is larger than one molecular layer, being able to produce the amount of light scattering responsible for the observed decrease. In fact, the pre-wetting layer connecting the ordered surface layer with the isotropic bulk is responsible for the observed scattering. At T_{IN}^{b} , the transmission strongly decreases, whereas a few degrees below T_{IN}^{b} , the sample completely recovers its transparency and stays transparent until room temperature is reached. Removal of the magnetic field at room temperature leaves both birefringence and transmission signals unchanged. Further optical characterizations show a planar, uniaxial monodomain alignment, with the director along the magnetic-field direction. After being aligned in a magnetic field, the sample was heated in the absence of an applied magnetic field and its relaxation was recorded. Upon heating, the birefringence follows a different path, and the bulk and the surface IN transition are no longer distinctly observed (Fig. 4a). Throughout the heating process the sample transmission maintains a constant maximum value (Fig. 4b). The dip in the optical transmission upon cooling always appears at T_{IN}^{b} and has the same width of 11.5 °C, independently of the substrate or the applied magnetic field. We associate this behaviour with an isotropic/nematic coexistence region, on the basis of the fact that E63 is a mixture of various LCs, which causes light scattering.

Some of the samples were cooled only in the partial presence of an applied magnetic field. For the samples that had the field applied below the bulk transition, orientation is observed but when the field is turned off the orientation relaxes. The same results were recorded for the samples that had the magnetic field applied above the bulk transition but below the temperature where the surface transition was complete. In contrast, the cells that had the field on until below the surface transition and then removed showed uniform stable alignment at room temperature, similar to the cells cooled in a field all the way from the isotropic phase down to room temperature. These control experiments prove that $\Delta b_{\text{IN}}^{\text{s}}$ is indeed related to a

surface transition and that the orientation of the surface layer is locked at the temperature corresponding to the bulk transition. Thus the ordered surface layer imposes its orientation on the bulk, being responsible for the alignment stability after the field is turned off.

To analyse these observations we start from the free energy density expression^{19,20} of a single-component nematic LC cell in a magnetic field, H :

$$F = f(Q, T) + L(dQ/dr)^2 + G\delta(z)Q \quad (1)$$

where $f(Q, T) = a(T - T^*)Q^2 - BQ^3 + CQ^4 + \Delta\chi QH^2$. Here Q , T and T^* are the uniaxial order parameter, temperature and minimum supercooling temperature respectively. G is the surface field, a , B and C are LC-specific constants, L is related to the elastic LC constants and $\Delta\chi$ is the anisotropy of the diamagnetic susceptibility of the LC molecules. Typical numbers confirmed by experiments^{17,21} show that the direct diamagnetic contribution $\Delta\chi QH^2$ is very small: an applied field of 10 T would only cause a change in the transition temperature of about 10 mK. Therefore in the following we will neglect this diamagnetic effect. The elastic term can be rewritten as $L(dQ/dr)^2 \approx LQ^2[A_D(H)/\xi]^2$ with $A_D(H)$ the mean amplitude of the director fluctuation and ξ the average correlation length. Using²² $\xi = [L/a(T - T^*)]^{-1/2}$ the free energy expression becomes:

$$F \approx a(T - T^*)Q^2[1 + A_D^2(H)] - BQ^3 + CQ^4 + G\delta(z)Q \quad (2)$$

Equation (2) shows that the nucleation of the nematic phase is more likely to appear on the surface. $A_D^2(H)$ varies between $A_D^2 \approx 0.1$ for no induced order and $A_D^2 = 0$ for perfect order, but the bulk starts to grow from the surface nuclei before the surface layer is complete, rendering the fluctuation effect much weaker. This, combined with T^* being only a few tens of a kelvin below T_{IN} (for single component LC), means that the bulk transition is completed over a narrow temperature range with little influence from the magnetic field.

However, for a multi-component LC mixture, the substrate can induce a small compositional drift near the surface, which subsequently leads to a complete separation of the surface and bulk transition, as shown in Fig. 2. For the temperature interval $T_{\text{IN}}^{\text{b}} < T < T_{\text{IN}}^{\text{s}}$, upon cooling, the ordered wetting layer tries to impose its orientation on the bulk. The free energy expression of the bulk component becomes:

$$F \approx a(T - T^*)Q^2[1 + \xi_{\text{bulk}}^2[A_D(H)/\xi_{\text{surface}}]^2] - BQ^3 + CQ^4 \quad (3)$$

The elastic energy due to the surface order fluctuation acts as an energetic barrier for the bulk component, causing a temperature gap between bulk and surface transitions, possible because for LC mixtures T^* can be much lower than T_{IN} . This is supported by the particular shape of the birefringence curves recorded upon cooling (Fig. 4a). For reasons similar to those described above, T_{IN}^{s} is not affected by the magnetic field. However, the field induces an alignment of the surface wetting layer below T_{IN}^{s} , which tends to suppress the surface order parameter fluctuations. It is this surface-bulk interaction term that is implicitly responsible for the field dependence of T_{IN}^{b} , as experimentally shown in Fig. 3.

When approaching the IN transition from below, the surface-bulk interaction term does not oppose the relaxation and Q follows the typical temperature relaxation path with the director locked by the surface orientation, as shown in Fig. 4a. For all cases we consistently find that the cells undergo the IN transition at 94.7 °C regardless of their different substrates and we cannot resolve the surface transition as separate.

The present results, showing the occurrence of a surface IN phase transition in an LC with its order dependent on the surface properties, are consistent with the predictions of refs 3 and 4, namely that the surface transition is of first order on surfaces with weak anchoring and is continuous in the presence of strong anchoring. □

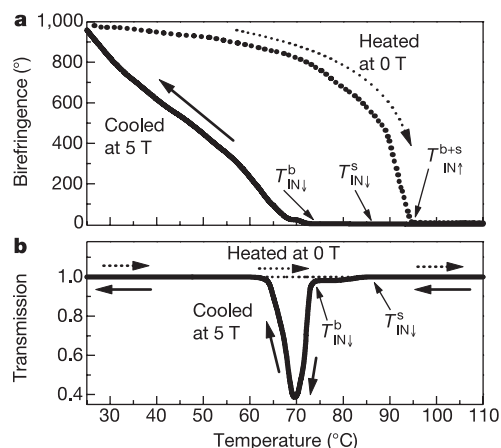


Figure 4 Surface-stabilized magnetic field induced alignment. Birefringence (a) and optical transmission (b) of a PVFM-coated cell upon cooling down in a field of 5 T (continuous line) and subsequently warming up in no applied field (dotted line). Subscript arrows denote the surface, T_{IN}^{s} , and the bulk, T_{IN}^{b} , transition temperatures upon cooling down and warming up, respectively.

Received 14 June; accepted 26 November 2002; doi:10.1038/nature01331.

1. Cahn, W. J. Critical point wetting. *J. Chem. Phys.* **66**, 3667–3670 (1977).
2. Nakanishi, H. & Fisher, M. E. Multicriticality of wetting, prewetting, and surface transitions. *Phys. Rev. Lett.* **49**, 1565–1568 (1982).
3. Sheng, P. Phase transition in surface-aligned nematic films. *Phys. Rev. Lett.* **37**, 1059–1062 (1976).
4. Sluckin, T. J. & Poniewierski, A. Novel surface phase transition in nematic liquid crystals. *Phys. Rev. Lett.* **55**, 2907–2910 (1985).
5. Yokohama, H. *et al.* Boundary dependence of the formation of new phase at the isotropic–nematic transition. *Mol. Cryst. Liq. Cryst.* **99**, 39–52 (1983).
6. Clark, N. A. Surface memory effects in liquid crystals: Influence of surface composition. *Phys. Rev. Lett.* **55**, 292–295 (1985).
7. Chen, W. *et al.* Orientational wetting behavior of a liquid crystal homologous series. *Phys. Rev. Lett.* **62**, 1860–1863 (1989).
8. Miyano, K. Wall-induced pretransitional birefringence: a new tool to study boundary aligning forces in liquid crystals. *Phys. Rev. Lett.* **43**, 51–54 (1979).
9. Hsiung, H. *et al.* Wall induced orientational order of a liquid crystal in the isotropic phase—an evanescent wave ellipsometry study. *Phys. Rev. Lett.* **57**, 3065–3068 (1986).
10. De Schrijver, P. *et al.* Surface-induced pretransitional order in the isotropic phase near the isotropic–nematic phase transition. *Liq. Cryst.* **21**, 745–749 (1996).
11. Crawford, G. P. *et al.* Ordering and self-diffusion in the 1st molecular layer at a liquid-crystal polymer interface. *Phys. Rev. Lett.* **66**, 723–726 (1991).
12. Akiyama, H. *et al.* Thermal stability of magnetically-aligned nematic liquid crystal layer on nonrubbed polyimide surface. *Jpn. J. Appl. Phys.* **36**, L1204–L1206 (1997).
13. Jerome, R. Surface effects and anchoring in liquid-crystals. *Rep. Prog. Phys.* **54**, 391–451 (1991).
14. Bonn, D. & Ross, D. Wetting transitions. *Rep. Prog. Phys.* **64**, 1085–1163 (2001).
15. van Haaren, J. Wiping out dirty displays. *Nature* **411**, 29–30 (2001).
16. Chaudhari, P. *et al.* Atomic-beam alignment of inorganic materials for liquid-crystal displays. *Nature* **411**, 56–59 (2001).
17. Maret, G. & Dransfeld, K. *Strong and Ultrastrong Magnetic Fields and Their Applications* (ed. Herlach, F.) 143–204 (Springer, Berlin, 1985).
18. Jakli, A. *et al.* Rotational viscosities of polymer solutions in a low molecular weight nematic liquid crystal. *Mol. Cryst. Liq. Cryst.* **198**, 331–340 (1991).
19. de Gennes, P. G. & Prost, J. *The Physics of Liquid Crystals*, 2nd edn (Clarendon, Oxford, 1993).
20. Sheng, P. *et al.* Disordered-surface-layer transition in nematic liquid crystals. *Phys. Rev. A* **46**, 946–950 (1992).
21. Rosenblatt, C. Magnetic field dependence of the nematic–isotropic transition temperature. *Phys. Rev. A* **24**, 2236–2240 (1981).
22. Singh, S. Phase transitions in liquid crystals. *Phys. Rep.* **34**, 107–269 (2000).

Acknowledgements Part of this work was supported by the EU-TMR Network SILC and the Dutch Technology Foundation STW.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.I.B. (e-mail: boamfa@sci.kun.nl).

Magnitude and timing of temperature change in the Indo-Pacific warm pool during deglaciation

Katherine Visser*, Robert Thunell* & Lowell Stott†

* Department of Geological Sciences, University of South Carolina, Columbia, South Carolina 29205, USA

† Department of Earth Sciences, University of Southern California, Los Angeles, California 90089, USA

Ocean–atmosphere interactions in the tropical Pacific region have a strong influence on global heat and water vapour transport and thus constitute an important component of the climate system^{1,2}. Changes in sea surface temperatures and convection in the tropical Indo-Pacific region are thought to be responsible for the interannual to decadal climate variability observed in extra-tropical regions^{1,3}, but the role of the tropics in climate changes on millennial and orbital timescales is less clear. Here we analyse oxygen isotopes and Mg/Ca ratios of foraminiferal shells from the Makassar strait in the heart of the Indo-Pacific warm pool, to obtain synchronous estimates of sea surface temperatures and

ice volume. We find that sea surface temperatures increased by 3.5–4.0 °C during the last two glacial–interglacial transitions, synchronous with the global increase in atmospheric CO₂ and Antarctic warming, but the temperature increase occurred 2,000–3,000 years before the Northern Hemisphere ice sheets melted. Our observations suggest that the tropical Pacific region plays an important role in driving glacial–interglacial cycles, possibly through a system similar to how El Niño/Southern Oscillation regulates the poleward flux of heat and water vapour.

Much research in the past two decades has pointed to the North Atlantic, and its prominence in controlling thermohaline circulation, as being the key to the driving of global climate variability on both millennial and orbital timescales⁴. The small (1–2 °C) changes in sea surface temperature (SST) predicted for the tropics for the Last Glacial Maximum (LGM) by the CLIMAP⁵ study also focused attention away from the tropics to the North Atlantic. However, recent studies suggest that tropical Pacific SSTs were at least 3 °C (ref. 6) and perhaps 5 °C (ref. 7) cooler during the LGM. Such changes in tropical SST should have a significant effect on the export of heat and water vapour to high latitudes, as well as atmospheric CO₂ concentrations⁸. Indeed, coupled ocean–atmosphere model studies indicate that heat flux from the tropics may be the most viable source of forcing for global climate change on both millennial and orbital timescales⁹.

Establishing the timing of climate change in the tropics relative to changes in other regions is necessary in order to evaluate the role that this region plays in driving global climate change. It has been suggested that changes in atmospheric CO₂ and Southern Hemisphere temperature are synchronous, and lead Northern Hemisphere ice volume changes by several thousand years^{10,11}. Conversely, it has been reported that the warming associated with the last deglaciation was synchronous between the two hemispheres¹².

We performed Mg/Ca and oxygen isotope analyses on the surface dwelling planktonic foraminifer *Globigerinoides ruber* from core MD9821-62 (4° 41.33' S, 117° 54.17' E; 1,855 m water depth) collected from the Makassar strait in the heart of the Indo-Pacific warm pool (IPWP; Fig. 1). This is a region of deep atmospheric convection, and the primary source of heat and water vapour flux to the atmosphere.

The δ¹⁸O of *G. ruber* is a recorder of the combined effects of seawater δ¹⁸O variability and the temperature at which the foraminifer calcite precipitates. The ¹⁸O composition of surface waters in the western tropical Pacific records both the effects of ice volume change and local salinity variability. The Mg/Ca ratio of planktonic foraminifers is a relatively new proxy for estimating SST. The

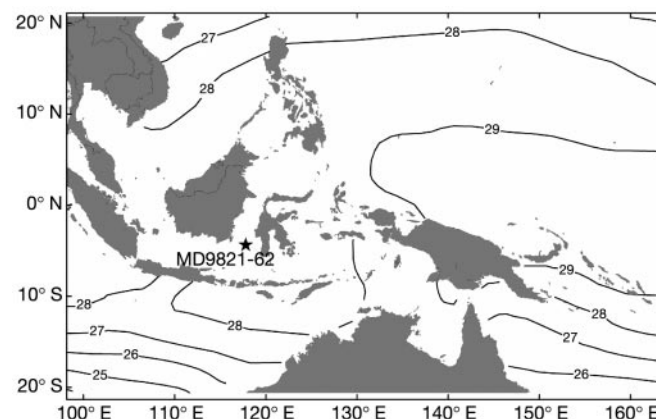


Figure 1 Map of sea surface temperature of the western Pacific and Indonesian region. The 'warm pool' is demarcated by the 28 °C isotherm. Core MD9821-62 was collected from the Makassar strait, Indonesia (4° 41.33' S, 117° 54.17' E, 1,855 m water depth).

temperature dependence of Mg incorporation into planktonic foraminiferal calcite has been verified both in culturing experiments¹³ and field studies¹⁴. As the Mg/Ca of planktonic foraminifer is not influenced by the same parameters as $\delta^{18}\text{O}$, it can be used in parallel with $\delta^{18}\text{O}$ to estimate what fraction of the $\delta^{18}\text{O}$ signal is due to ice volume or local salinity changes. Additionally, by measuring Mg/Ca and $\delta^{18}\text{O}$ on the same samples we can determine the relative phasing between tropical SST change and high latitude deglaciation. To calculate SST, we used an equation¹⁵ developed for *G. ruber* from South China Sea surface sediment samples recovered from depths less than 2,000 m: $\text{Mg/Ca (mmol mol}^{-1}) = 0.38 \exp[0.089 \text{ SST (}^{\circ}\text{C)}]$.

This equation has same exponential term as the equation developed by Lea *et al.*⁶ for the equatorial Pacific, but differs by having a larger pre-exponential constant. It has been shown that increased dissolution decreases the pre-exponential constant¹⁶. This suggests that the South China Sea samples used in the Hastings *et al.*¹⁵ calibration may have been better preserved than the tropical Pacific samples used by Lea *et al.*⁶. Several lines of evidence indicate that calcite is well preserved in our core. First, the core depth (1,855 m) is well above both the present day ($\sim 3,400$ m) and glacial-age lysocline depths ($\sim 4,100$ m)¹⁷. Second, aragonite, which dissolves more easily than calcite, is present throughout the core. Third, there is no significant difference in the average shell weight of *G. ruber* between the interglacial (11.6 and 11.1 μg per shell for marine isotope stages (MIS) 1 and 5e, respectively) and glacial periods (11.3 and 11.8 μg per shell for MIS 2 and 6, respectively). Thus, we chose to use the Hastings *et al.*¹⁵ equation, because our core is relatively unaffected by dissolution and the estimated Holocene temperatures are very similar to modern day temperatures (~ 28 – 29°C) at our core location.

The age model for the portion of the core that extends back

through the last glacial is based on 10 accelerator mass spectrometry (AMS) ^{14}C dates that were determined on *G. ruber* and converted to calendar years. For the penultimate glacial and last interglacial (MIS 6 and 5e), the *G. ruber* $\delta^{18}\text{O}$ record was correlated to the SPECMAP stacked $\delta^{18}\text{O}$ chronology¹⁸. The time scale was constructed assuming constant sedimentation rates between radiocarbon dates and isotope tie points. On the basis of this age model, the average sample spacing is ~ 400 yr.

We used two different approaches to estimate the magnitude of the temperature changes across the last two glacial terminations from the Mg/Ca data. First, we averaged the temperatures determined for all samples within each of the glacial (MIS 2 and 6) and interglacial (MIS 1 and 5e) intervals (Fig. 2). This approach yields SST changes of $3.3 \pm 0.6^{\circ}\text{C}$ and $4.1 \pm 0.6^{\circ}\text{C}$ across terminations I and II, respectively, and are considered conservative estimates of the total SST change. These data also suggest that, on average, western tropical Pacific SSTs were the same during stages 2 and 6 (25.8 and 25.9°C), and that the last interglacial was approximately 1°C warmer than the Holocene (30.0 versus 29.1°C) at this location. The second approach uses three consecutive samples that represent the minimum glacial and maximum interglacial temperatures, and yields temperature changes of $3.7 \pm 0.6^{\circ}\text{C}$ and $5.3 \pm 0.7^{\circ}\text{C}$ across terminations I and II, respectively (Fig. 2). These temperature estimates are supported by our oxygen isotope data. The amplitudes of the $\delta^{18}\text{O}$ changes across terminations I and II are 1.7‰ and 1.9‰, respectively (Fig. 2), of which 0.9‰ and 1.1‰, respectively, is attributed to the change in global ice volume^{11,19}. Assuming no salinity change, this translates into a temperature change of $\sim 3.6^{\circ}\text{C}$ across both terminations, using the $\delta^{18}\text{O}$ –temperature relationship of $0.22\text{‰ } \delta^{18}\text{O} \equiv 1^{\circ}\text{C}$ (ref. 20). These Mg/Ca and $\delta^{18}\text{O}$ derived temperature changes are larger than those estimated by CLIMAP⁵ for this region, but are in line with a growing body of evidence that

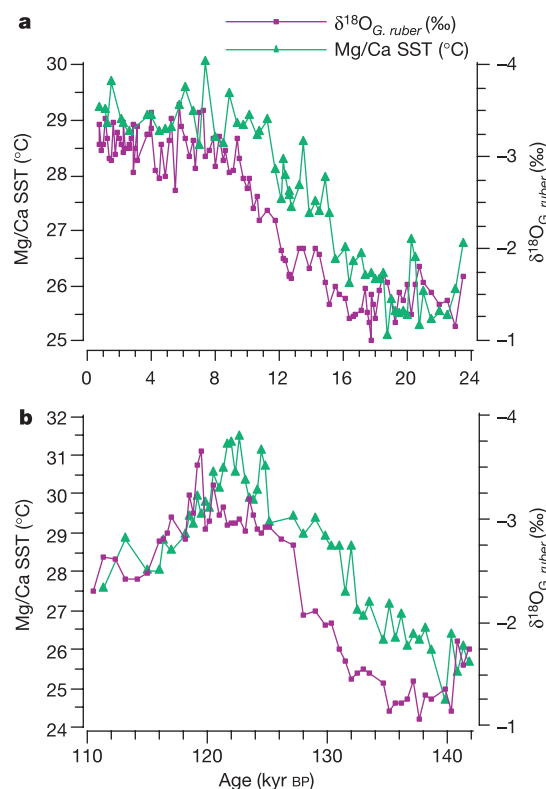


Figure 2 *Globigerinoides ruber* oxygen isotope (purple curve) and Mg/Ca SST (green curve) records for core MD9821-62 plotted versus time for MIS 1 and 2 (a) and MIS 5e and 6 (b). For both terminations, the surface temperature warming occurs several thousand years before the decrease in ice volume reflected by the $\delta^{18}\text{O}$ record.

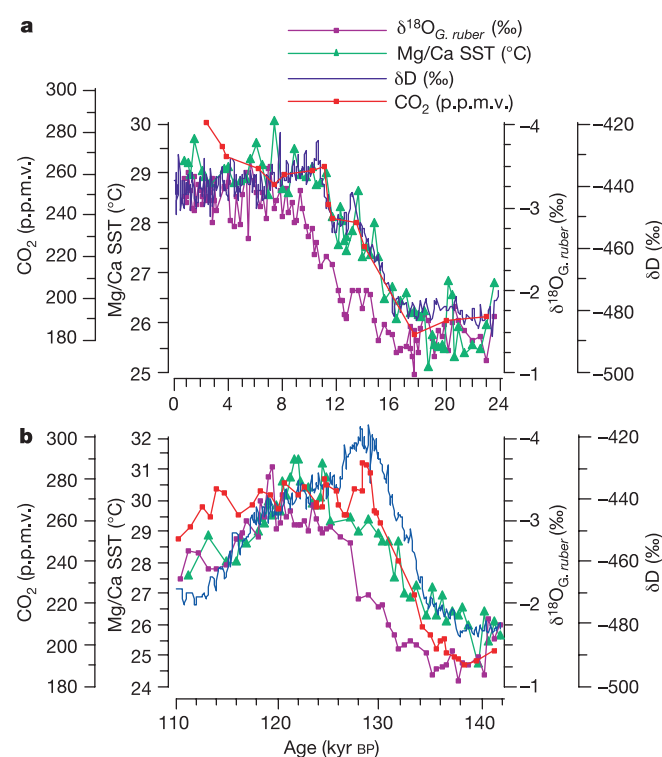


Figure 3 Comparison of the MD9821-62 *G. ruber* $\delta^{18}\text{O}$ (purple curve) and Mg/Ca SST (green curve) records with the Vostok CO_2 (red curve) and deuterium isotope (blue curve) records²². The deuterium record reflects changes in atmospheric temperature over Antarctica. The change in the Mg/Ca, deuterium and CO_2 records during both deglaciations occurs before the $\delta^{18}\text{O}$ change.

suggests that the tropical oceans were 3–5 °C colder during the LGM^{6,7}.

For both terminations, the change in Mg/Ca is not synchronous with the change in $\delta^{18}\text{O}$ (Fig. 2). As the *G. ruber* $\delta^{18}\text{O}$ record reflects both local changes in SST and global ice volume, the timing of change in the Mg/Ca and $\delta^{18}\text{O}$ records should be similar during the period of surface temperature warming associated with deglaciation. However, for both terminations, the $\delta^{18}\text{O}$ records continue to decrease for several thousand years after the Mg/Ca records indicate that SST reached interglacial values. This additional change in $\delta^{18}\text{O}$ is interpreted as reflecting the decrease in global ice volume. Within the limits of our age model, we estimate that the $\delta^{18}\text{O}$ ice volume changes lag the Mg/Ca SST changes by 2,000–3,000 yr. For example, during termination I, SST reaches the modern interglacial value of 29 °C by ~12 kyr ago, while $\delta^{18}\text{O}$ continues to decrease until at least 10 kyr ago. Our findings indicate that the tropical western Pacific region warmed before the melting of Northern Hemisphere ice sheets during the last two deglaciations. This phase relationship is consistent with that recently reported for the equatorial Pacific⁶.

What are the broader implications of our findings concerning the magnitude and timing of SST change in the Indo-Pacific warm pool during the last two glacial–interglacial transitions? The global atmosphere is very sensitive to changes in tropical SST², particularly with regards to its effect on fluxes of heat and water vapour to higher latitudes. The tropics, in general, and the Indo-Pacific warm pool, in particular, are the main regions from which water vapour is supplied to the atmosphere. As water vapour is one of the main greenhouse gases, substantial variations in the moisture content of the atmosphere associated with past changes in SST can be expected to have had a major role in global climate change. It has been estimated that during the last glacial period the absolute water vapour content of the atmosphere at sea level in the tropics was 70% of the modern value, and that the greenhouse effect of this decrease in water vapour led to a global cooling of 2 °C (ref. 21). Similarly, a decrease in tropical Indo-Pacific SST during glacial periods of 3.5–5 °C should contribute to the 80 p.p.m. decrease in atmospheric CO₂ concentration recorded in ice cores²², owing to the fact that CO₂ is more soluble in colder water. For a 1 °C increase in temperature, the equilibrium partial pressure of CO₂ exerted by sea water increases by ~11–16 p.p.m. (ref. 8). Therefore, a 3.5 °C cooling in the tropics could account for a substantial portion of the atmospheric decrease in CO₂ during glacial periods, suggesting that the low latitude oceans have a significant effect on atmospheric CO₂ concentrations. Indeed, the changes in atmospheric CO₂ are synchronous with our estimated SST changes in the Indo-Pacific for the last two terminations (Fig. 3).

The fact that SST change leads the ice volume change by 2,000–3,000 yr during the past two terminations at our study site supports the notion that the tropical Pacific is actively involved in forcing global climate change^{1,2,9}. As noted previously⁶, the increase in SST in the tropical Pacific during deglaciation is synchronous with an increase in the Vostok ice-core deuterium isotope record²², which reflects atmospheric warming over Antarctica (Fig. 3). The present day pattern of meridional heat transport by the oceans and atmosphere may provide some insight into the mechanism that links climate change in the tropical Pacific and the high southern latitudes. At present, the patterns of surface heat transport are quite different between the Atlantic and Pacific oceans. In the Atlantic, heat moves northward at all latitudes and results in cross-equatorial heat transport. By contrast, the tropical Pacific transports heat to both hemispheres, although this export is asymmetrical in that roughly twice as much heat is directed to the south than the north²³.

We suggest that past changes in the transport of heat to the high southern latitudes may have been regulated by a system analogous to the El Niño/Southern Oscillation (ENSO). Present day variability in SST, sea-ice extent, and wind stress in the Antarctic region is tied, via atmospheric teleconnections, to tropical Pacific ENSO activity²⁴.

Recent work suggests that the tropical Pacific was dominated by an El Niño-like climate during the LGM²⁵. This would have decreased convection in the western Pacific, and reduced both zonal and meridional circulation. Conversely, a switch to La Niña-like conditions occurred during deglaciation^{25,26}, and reinvigorated heat transport to the high southern latitudes.

Additionally, the warming of surface waters in these regions during deglaciation would have resulted in a flux of CO₂ to the atmosphere, which would have had a positive feedback on global warming. A new view of the importance of the tropics in controlling global climate change is beginning to emerge, although the nature of the linkage between tropical and extra-tropical regions still needs to be resolved. □

Methods

Oxygen isotope analyses

These were conducted on specimens (60–70 µg) of the planktonic foraminifer *G. ruber* picked from the 250–355 µm size fraction. Analyses were carried out using a VG Optima stable isotope ratio mass spectrometer (IRMS) equipped with an automated carbonate system. The samples were reacted in a common acid bath of 100% phosphoric acid at 90 °C. The evolved CO₂ gas was then analysed on the IRMS. Oxygen and carbon isotopic data are reported in delta notation relative to (Vienna) Pee Dee Belemnite. The long-term standard reproducibility for oxygen ($\delta^{18}\text{O}$) isotopes based on replicate measurements of a reference standard is $\pm 0.07\text{‰}$.

Mg/Ca analyses

The Mg/Ca ratio in *G. ruber* was determined on a Jobin-Yvon Ultima inductively coupled plasma-atomic emission spectrophotometer (ICP-AES). For this study, the two lines used were: Ca (317.93 nm wavelengths) and Mg (285.213 nm). Approximately 220–240 µg (~23 shells from the 250–355 µm size fraction) of pre-cleaned (in methanol) whole foraminifers were processed using the following procedure first developed by Boyle²⁷. First, the samples were gently crushed in acid-cleaned 1.5-ml microcentrifuge tubes, making sure that all chambers were fully open. In order to remove the fine clays, the foraminifer fragments were covered with deionized water (DI H₂O), sonicated for ~1 min, resuspended by the addition of 400 µl DI H₂O; then the supernatant was siphoned off using a micropipette. These steps were repeated twice with DI H₂O, twice with methanol, and then once with DI H₂O to remove the methanol. To remove organic matter, 250 µl of an oxidizing reagent (0.1 N NaOH and 0.15% H₂O₂) was added. The samples were capped and placed in a boiling water bath for 5 min, sonicated for 1 min, then replaced in the boiling water bath for 5 min. Afterwards, the oxidizing solution was removed, and the samples were rinsed twice with 400 µl of DI H₂O to ensure complete removal of the oxidizing agent. The samples were then transferred to new, acid-cleaned microcentrifuge tubes. The cleaned shells were leached with weak acid (0.001 M HNO₃) and rinsed with DI H₂O. Finally, ~500 µl of 5% HNO₃ was added to dissolve the foraminifer shells. It is estimated that ~40–60% of the initial sample weight was lost through the cleaning procedure. The *G. ruber* Mg/Ca equation developed by Hastings *et al.*¹⁵ has the same exponential and pre-exponential constants as the equation recently presented by Dekens *et al.*²⁸ The difference between the two equations lies in the fact that the latter equation has an added term for water depth. The Hastings *et al.*¹⁵ equation is used in this study because it yields Holocene temperatures that are consistent with present day temperatures at our core location. The standard error of estimate for the various temperature equations derived from core top calibrations is typically in the range of 0.5–1.0 °C (refs 6, 14, 15, 28).

Calibration of radiocarbon dates

AMS ¹⁴C dates were determined on monospecific samples (4–10 mg) of the planktonic foraminifer *G. ruber* at the Center for Accelerator Mass Spectrometry, Lawrence Livermore National Laboratory. The youngest eight ¹⁴C ages were converted to calendar ages using the calibration program CALIB 4.3 (ref. 29) and the oldest two ¹⁴C ages were converted using the equation $(-698.65 \times 10^{-3} + 1.2695(^{14}\text{C}) - [2.7039 \times 10^{-6} (^{14}\text{C})^2])$ (we find $r^2 = 0.999$; standard error = ± 150 yr), where ¹⁴C is the reservoir-corrected ¹⁴C age³⁰. In order to account for the difference in ¹⁴C composition between the atmosphere and the surface ocean, a reservoir correction of 400 yr was subtracted from the ¹⁴C dates before converting them to calendar years³¹.

Received 7 May; accepted 5 November 2002; doi:10.1038/nature01297.

1. Cane, M. A role for the tropical Pacific. *Science* **282**, 59–61 (1998).
2. Pierrehumbert, R. Climate change and the tropical Pacific: The sleeping dragon wakes. *Proc. Natl Acad. Sci.* **97**, 1355–1358 (2000).
3. Hoerling, M., Hurrell, J. & Xu, T. Tropical origins for recent North Atlantic climate change. *Science* **292**, 90–92 (2001).
4. Broecker, W. Paleocene circulation during the last deglaciation: a bipolar seesaw? *Paleoceanography* **13**, 119–121 (1998).
5. CLIMAP Project Members. Seasonal reconstructions of the Earth's surface at the last glacial maximum. *Geol. Soc. Am. Map Chart Ser.* **MC-36**, 1–18 (1981).
6. Lea, D., Pak, D. & Spero, H. Climate impact of late Quaternary equatorial Pacific sea surface temperature variations. *Science* **289**, 1719–1724 (2000).
7. Guilderson, T., Fairbanks, R. & Rubenstein, J. Tropical Atlantic coral oxygen isotopes; glacial-interglacial sea surface temperatures and climate change. *Mar. Geol.* **172**, 75–89 (2001).

8. Bacastow, R. The effect of temperature change of the warm surface waters of the oceans on atmospheric CO₂. *Glob. Biogeochem. Cycles* **10**, 319–333 (1996).
9. Cane, M. & Clement, A. *Mechanisms of Global Climate Change at Millennial Time Scales* AGU Geophysical Monograph 112 (eds Clark, P., Webb, R. & Keigwin, L.) 373–383 (American Geophysical Union, Washington DC, 1999).
10. Broecker, W. & Henderson, G. The sequence of events surrounding Termination II and their implications for the cause of glacial-interglacial CO₂ changes. *Paleoceanography* **13**, 352–364 (1998).
11. Shackleton, N. The 100,000-year ice age cycle identified and found to lag temperature, carbon dioxide, and orbital eccentricity. *Science* **289**, 1897–1902 (2000).
12. Steig, E. *et al.* Synchronous climate changes in Antarctica and the North Atlantic. *Science* **282**, 92–95 (1998).
13. Lea, D., Mashiotta, T. & Spero, H. Controls on magnesium and strontium uptake in planktonic foraminifer determined by live culture. *Geochim. Cosmochim. Acta* **63**, 2369–2379 (1999).
14. Elderfield, H. & Ganssen, G. Past temperature and $\delta^{18}\text{O}$ of surface ocean waters inferred from foraminiferal Mg/Ca ratios. *Nature* **405**, 442–445 (2000).
15. Hastings, D., Kienast, M., Steinke, S. & Whitko, A. A Comparison of three independent paleotemperature estimates from a high resolution record of deglacial SST records in the tropical South China Sea. *Eos* **82**, PP12B-10 (2001).
16. Rosenthal, Y., Lohmann, G., Lohmann, K. & Sherrell, R. Incorporation and preservation of Mg in *Globigerinoides sacculifer*: Implications for reconstructing the temperature and $^{18}\text{O}/^{16}\text{O}$ of seawater. *Paleoceanography* **15**, 135–145 (2000).
17. Farrell, J. & Prell, W. Climatic change and CaCO₃ preservation: An 800,000 year bathymetric reconstruction from the central equatorial Pacific Ocean. *Paleoceanography* **4**, 447–466 (1989).
18. Martinson, D. *et al.* Age dating and the orbital theory of the ice-ages: development of a high-resolution 0–300,000 year chronostratigraphy. *Quat. Res.* **27**, 1–29 (1987).
19. Schrag, D., Hampt, G. & Murray, D. Pore fluid constraints on the temperature and oxygen isotopic composition of the glacial ocean. *Science* **272**, 1930–1932 (1996).
20. Epstein, S., Buchsbaum, R., Lowenstam, H. & Urey, H. Revised carbonate-water isotopic temperature scale. *Geol. Soc. Am. Bull.* **64**, 1315–1325 (1953).
21. Broecker, W. Mountain glaciers: Records of atmospheric water vapor content. *Glob. Biogeochem. Cycles* **11**, 589–597 (1997).
22. Petit, J. *et al.* Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature* **399**, 429–436 (1999).
23. Hastenrath, S. On meridional heat transports in the world ocean. *J. Phys. Ocean.* **12**, 922–927 (1982).
24. White, W. B. & Peterson, R. An Antarctic circumpolar wave in surface pressure, wind, temperature and sea-ice extent. *Nature* **380**, 699–702 (1996).
25. Koutavas, A., Lybch-Stieglitz, J., Marchitto, T. M. & Sachs, J. El Niño-like pattern in ice age tropical Pacific sea surface temperature. *Science* **297**, 226–230 (2002).
26. Stott, L., Poulsen, C., Lund, S. & Thunell, R. Super ENSO and global climate oscillations at millennial time scales. *Science* **297**, 222–226 (2002).
27. Boyle, E., Labeyrie, L. & Duplessy, J.-C. Calcitic foraminiferal data confirmed by cadmium in aragonitic *Hoeglundina*; application to the last glacial maximum in the northern Indian Ocean. *Paleoceanography* **10**, 881–900 (1995).
28. Dekens, P. S., Lea, D., Pak, D. & Spero, H. Core top calibration of Mg/Ca in tropical foraminifer: Refining paleotemperature estimation. *Geochim. Geophys. Res.* **3**, 10.1029/2001GC000200 (2002).
29. Stuiver, M. *et al.* INTCAL98 Radiocarbon age calibration 24,000–0 cal BP. *Radiocarbon* **40**, 1041–1083 (1998).
30. Guilderson, T., Burckle, L., Hemming, S. & Peltier, W. Late Pleistocene sea level variations derived from the Argentine Shelf. *Geochim. Geophys. Res.* **1**, 10.1029/2000GC000098 (2000).
31. Bard, E. Correction of accelerator mass spectrometry ^{14}C ages measured in planktonic foraminifer: Paleoceanographic implications. *Paleoceanography* **3**, 635–645 (1988).

Acknowledgements We thank E. Tappa for technical assistance. This work was supported by the US National Science Foundation (R.T. and L.S.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.T. (e-mail: thunell@geol.sc.edu)

Group decision-making in animals

L. Conradt & T. J. Roper

School of Biological Sciences, University of Sussex, Brighton BN1 9QG, UK

Groups of animals often need to make communal decisions, for example about which activities to perform¹, when to perform them^{2–9} and which direction to travel in^{1,6,7}; however, little is known about how they do so^{10–12}. Here, we model the fitness consequences of two possible decision-making mechanisms: ‘despotism’^{6,7,10} and ‘democracy’^{1,6,7,10}. We show that under most conditions, the costs to subordinate group members, and to the group as a whole, are considerably higher for despotic than for

democratic decisions. Even when the despot is the most experienced group member, it only pays other members to accept its decision when group size is small and the difference in information is large. Democratic decisions are more beneficial primarily because they tend to produce less extreme decisions, rather than because each individual has an influence on the decision *per se*. Our model suggests that democracy should be widespread and makes quantitative, testable predictions about group decision-making in non-humans.

Notwithstanding extensive literature on decision-making by animals acting alone^{13–15}, group decision-making processes have been largely neglected from a theoretical point of view^{6,7,10}. Two extreme mechanisms whereby a group could in principle reach communal decisions are: (1) despotically, where one dominant decides; and (2) democratically, where a majority of group members decides. The relative fitness consequences of these mechanisms are unknown. Many authors have assumed despotism without testing^{6,7,10}, because the feasibility of democracy, which requires the ability to vote and to count votes, is not immediately obvious in non-humans. However, empirical examples of ‘voting’ behaviours include the use of specific body postures^{1,10,11}, ritualized movements^{6,7,9,11,16}, and specific vocalizations^{5,8,12}, whereas ‘counting of votes’ includes adding-up to a majority of cast votes^{5–8,11}, integration of voting signals until an intensity threshold is reached^{9,16} and averaging over all votes^{1,12,16} (Table 1). Thus, democracy may exist in a range of taxa and does not require advanced cognitive capacity. Here, we model the fitness consequences of despotic and democratic decisions. One important context in which social animals frequently have to make communal decisions is in the duration of group activities^{1–7,17}. This ‘activity synchronization’ is essential if a group is to remain spatially coherent^{1–7,10}. However, to reach consensus decisions about the duration of activities, group members often have to compromise their own optimal activity budgets^{1,3,18,19}, the costs of which (synchronization costs) can be an important factor in shaping the social organization of populations^{1–4,17}. Our model compares the synchronization costs of despotic and democratic groups.

The model in Box 1 shows that synchronization costs are usually higher for despotic than for democratic groups. However, if synchronization costs are asymmetric (that is, stopping too late costs less than stopping too early, or vice versa), a modified democratic decision is least costly. In this case, the decision should be based not on a 50% majority, but on a proportion of members that depends on the degree of asymmetry in costs. For example, if stopping too early is twice as costly as stopping too late, the group

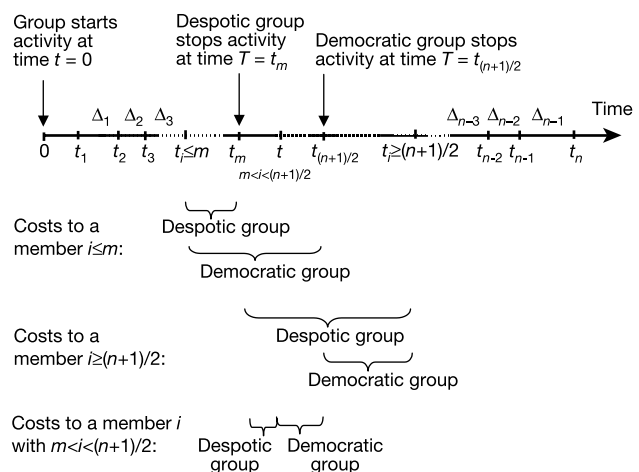


Figure 1 Scheme of a despotic or democratic decision about the duration of a single group activity session, and resulting costs. (See Box 1 for definitions.)

Box 1

Model of synchronization costs

A synchronized^{3–6,17} group of n individuals starts an activity session together at time $t = 0$. The group has to decide when to stop the activity. Each member has its own optimal stopping time^{1–4,19,22} $t_1, t_2, \dots, t_n$ (Fig. 1). A democratic group stops when a majority of members wants to stop; that is, at time $T = t_{(n+1)/2}$. A despotic group stops when the despot (member m) wants to stop (time $T = t_m$). Assuming that (1) synchronization costs increase linearly with the difference between a member's optimal and the group's realized activity duration; and (2) costs of stopping too late or too early are symmetrical. Then the costs to a member $i < (n+1)/2$ in a democratic group are:

$$C_i = c \sum_{j=1}^{(n-1)/2} \Delta_j \quad (1)$$

where $\Delta_i = t_{(i+1)} - t_i$ for $i < n$ (Fig. 1), and c is a constant 'cost per time'. The costs to members $i \geq (n+1)/2$ and to despotic members are analogous. Thus, the total costs to a democratic group (C_{dg}) are:

$$C_{dg} = c \left[\sum_{i=1}^{(n-1)/2} \sum_{j=1}^{(n-1)/2} \Delta_i + \sum_{i=(n+3)/2}^n \sum_{j=(n+1)/2}^{i-1} \Delta_i \right] \\ = c \left[\sum_{i=1}^{(n-1)/2} i \Delta_i + \sum_{i=(n+1)/2}^{n-1} (n-i) \Delta_i \right] \quad (2)$$

and to a despotic group (C_m):

$$C_m = c \left[\sum_{i=1}^{m-1} \sum_{j=1}^{m-1} \Delta_i + \sum_{i=m+1}^n \sum_{j=m}^{i-1} \Delta_i \right] = c \left[\sum_{i=1}^{m-1} i \Delta_i + \sum_{i=m}^{n-1} (n-i) \Delta_i \right] \quad (3)$$

The costs are higher in a despotic than a democratic group if:

$$c \sum_{i=m}^{(n-1)/2} (n-2i) \Delta_i > 0 \text{ for } m < (n+1)/2 \quad (4)$$

and analogously for $m \geq (n+1)/2$. As $(2i-n) > 0$ for $i \geq (n+1)/2$ and $(n-2i) > 0$ for $i \leq (n+1)/2$ (except if $m = (n+1)/2$), it follows that synchronization costs are always higher for despotic than democratic groups.

Relaxing assumption (1), costs are still higher for despotic than for democratic groups in most cases (see Supplementary Information for mathematical details and some exceptions). Relaxing assumption (2) (that is, cost per time $c_{\text{too early}} \neq c_{\text{too late}}$), a modified democratic decision is least costly, in which a proportion, $x = c_{\text{too early}} / (c_{\text{too late}} + c_{\text{too early}})$, of members decides instead of the usual 50% majority (see Supplementary Information).

should stop the activity when two-thirds of its members want to stop. Members that want to stop late should be twice as motivated to influence the decision as members that want to stop early, so integration of voting signals might be the best 'counting of votes' mechanism. Modified democratic decision-making mechanisms are comparable to the tradition, in many human societies, of using a two-thirds majority rather than a 50% majority for decisions that are potentially more costly if taken than if not taken (for example, constitutional changes in Germany²⁰).

The lower costs to a democratic group as a whole imply that democratic decisions are primarily beneficial because they lead to less extreme decisions than despotic decisions (Fig. 1), apart from the advantage that individual members obtain by influencing the decision in their favour. As synchronization costs can accumulate over consecutive sessions of group activities³, the difference in costs between despotic and democratic groups could become quite substantial^{1–4}.

The distribution of synchronization costs between group members depends on how homogeneous members are with respect to their optimal time budgets (Box 2). The despot, of course, always pays lower costs in a despotic group. In a homogeneous group, all other members benefit from a democratic decision. This is not true for extremely heterogeneous groups, but even here most of the members benefit from democracy.

No assumptions were made in our model about the means of enforcing a despotic decision (for example, coercion, manipulation or persuasion). However, it is obvious that a despot should not invest more energy than it can gain from despotism. Equally, subordinates should not invest more energy in resisting a despot than they can gain from a democratic decision. Our model suggests that, although single subordinates should not invest as much energy into resistance as the despot is prepared to invest into imposing the decision, all subordinates together should always invest more energy than the despot (Box 2). This suggests that despotic communal decisions might be rare. However, several like-minded members could join forces, make decisions between themselves democratically, but coerce the rest of the group. The result would be an oligarchy¹⁰ (for example, voting is often restricted to adults; Table 1).

Individuals might have incomplete information about their own optimal activity durations because they have imperfect knowledge about their environment (for example, the status of feeding patches, predation risks) and thus, do not exactly know the value to themselves of stopping an activity slightly earlier or later²¹. In

Table 1 Examples of democratic decisions in social animals

Decision	Species	Voting behaviour	Decision mechanism	N	Result
AC	Red deer (<i>Cervus elaphus</i>)	Standing up	Majority of adults decides	10	Group moves when mean 62% (s.d. 8%) of adults stand up*
AC	Gorilla (<i>Gorilla gorilla</i>)	Calling	Majority of adults decides	28	Group moves when median 65% (range 43–86%) of adults call ⁵
AC	Guinea baboons (<i>Papio papio</i>)	Movements	Majority decides	–	Anecdotal report ⁶
AC	Hamadryas baboons (<i>P. hamadryas</i>)	Movements	Majority decides	–	Anecdotal report ⁶
AC	Howler monkeys (<i>Alouatta palliata</i>)	Movements	Majority decides	–	Anecdotal report ⁷
AC	African elephant (<i>Loxodonta africanus</i>)	Low-frequency grumbles	Majority of adult females decides	–	Anecdotal report ⁸
AC	Whooper swans (<i>Cygnus cygnus</i>)	Head movements	Intensity of signals reaches threshold	54	Group flies when signalling intensity ≥ 26.7 signals min ⁻¹ (ref. 9)
DT	African buffalo (<i>Syncerus caffer</i>)	Direction of gaze	Mean of votes of adult females	13	Average angular difference between mean gazing direction and group travel direction = 3° (range 0–9°)*
DT	Hamadryas baboons (<i>P. hamadryas</i>)	Position on resting rock	Majority of adult males decides	155	In 131 of 155 observations the travel direction equalled the majority vote ¹¹
DT	Yellow baboons (<i>P. cynocephalus</i>)	Body orientation	Adults decide	–	Anecdotal report ¹⁰
DT	White-faced capuchins (<i>Cebus capucinus</i>)	Calls	Direction changes continuously with each caller	–	Anecdotal report ¹²
CN	Honeybees (<i>Apis mellifera</i>)	Dances	Integration of signals	–	A complex weighing of voting intensities and number of voters; authors provide a large data set in support ¹⁶

AC, activity change; CN, choice of nest site; DT, direction of travel; N, number of observations.

*L.C., unpublished data.

Box 2

Distribution of costs between members

In a homogeneous group⁴, each member (including the despot) is equally likely to be on any one occasion member 1, 2, ..., n who wants to stop the activity session 1st, 2nd, ..., n th. Thus, in a democratic group, the expected costs to each member are $1/n$ th of the total group costs (Box 1, equation (2)). In a despotic group, the expected group costs are $1/n \sum_{m=1}^n C_m$ (Box 1). None of these costs is paid by the despot, but they are equally shared by the $(n-1)$ subordinate members. Thus, a subordinate member benefits by a democratic decision, if:

$$\frac{2c}{n(n-1)} \sum_{i=1}^{n-1} i(n-i)\Delta_i > \frac{c}{n} \left[\sum_{i=1}^{(n-1)/2} i\Delta_i + \sum_{i=(n+1)/2}^{n-1} (n-i)\Delta_i \right]$$

$$\Rightarrow \sum_{i=1}^{(n-1)/2} \frac{n-2i+1}{n-1} i\Delta_i + \sum_{i=(n+1)/2}^{n-1} \frac{2i-n+1}{n-1} (n-i)\Delta_i > 0 \quad (5)$$

This inequality is always true (compare to Box 1, equation (4)).

In an extremely heterogeneous group^{4,22}, the same member is always member 1, 2, ..., or n who wants to stop 1st, 2nd, ..., or n th. The despot is always member m . Since the problem is symmetrical, we only consider the case $m < (n+1)/2$. All m members for which $t_i \leq t_m$ profit by a despotic decision (Fig. 1). However, all $(n+1)/2$ members for which $t_i \geq t_{(n+1)/2}$ profit by a democratic decision. The $[(n-1)/2 - m]$ members for which $t_m < t_i < t_{(n+1)/2}$ might or might not benefit by a democratic decision, depending on the values of $\Delta_m \dots \Delta_{(n-1)/2}$. Thus, most of the members (at least $(n+1)/2$ members) always profit by a democratic decision.

The distribution of costs between members can be used to predict the amounts of energy that a despot should invest in coercing a decision, and subordinates in resisting the despot. If costs to the despot plus costs to the subordinate in a democratic group are higher than costs to the subordinate in a despotic group, then the despot can invest more energy in coercion than the subordinate in resistance. This is usually the case (see Supplementary Information). However, all subordinates together can always invest more energy in resistance than the despot in coercion, because democratic group costs are always lower than despotic group costs (Box 1).

such a case, a group might be 'led' by its most experienced member^{6,7,10,21}. However, our model shows that it only pays group members to follow the decision of a more experienced leader, rather than to make the decision democratically, if errors are large relative to differences in optimal activity durations between members, group size is small, and the difference in experience between leader and ordinary group members is so large that the leader's average error is lower than the average median error of all other group members (Box 3).

The current shortage of studies of group decision-making processes in animals is due to a lack of testable, well-structured concepts and hypotheses relating to group decisions^{5-7,10,11,21}. Our model suggests that democratic decisions, being more beneficial than despotic decisions in most circumstances, should be widespread in animals. It also makes concise, testable and quantitative predictions about when groups should make decisions democratically; which proportion of group members should constitute the critical threshold for a decision, if costs are skewed; what the expected decision results are (for example, mean or median of voting members' requirements); which group members should try to leave democratic or despotic groups; how much members should invest into coercion or resistance to coercion of decisions; when to expect oligarchies; when to expect an experienced leader as decision maker; and which 'counting of votes' mechanism to expect in a particular case. □

Box 3

Groups with incomplete information

We assume that members make a random error, ε_i , when they determine their own optimal activity duration t_i , so that member i assumes its optimal activity duration to be $t'_i = t_i + \varepsilon_i$; where ε_i is normally distributed with mean 0 and variance σ_i^2 . One member (the potential leader; σ_l) is more experienced than other members (σ_{om}) thus $\sigma_l^2 < \sigma_{om}^2$. Decisions are made either despotically by the leader or democratically. Errors are either relatively small ($|\varepsilon_i| \ll \Delta_j$ for $i = 1, \dots, n, j = 1, \dots, n-1$), or large ($|\varepsilon_i| \gg \Delta_j$ for $i = 1, \dots, n, j = 1, \dots, n-1$). A despotic group stops the activity session at time $T = t'_l = t_l + \varepsilon_l$, and the expected group costs are (using equations (3) and (5)):

$$2c/n \sum_{i=1}^{n-1} i(n-i)\Delta_i + c/n \sum_{m=1}^n [(m-1)\varepsilon_l - (n-m)\varepsilon_l] + c|\varepsilon_l|$$

$$= 2c/n \sum_{i=1}^{n-1} i(n-i)\Delta_i + c|\varepsilon_l| \quad (6)$$

If errors are relatively small, the sequence with which members want to stop the activity is not different from the sequence of their optimal stopping times t_i . Thus, a democratic group stops at time $T = t'_{(n+1)/2} = t_{(n+1)/2} + \varepsilon_{(n+1)/2}$. A group would benefit from a leader decision if (using equations (2), (5) and (6)):

$$c/n \left[\sum_{i=1}^{(n-1)/2} (n-2i)i\Delta_i + \sum_{i=(n+1)/2}^{n-1} (2i-n)(n-i)\Delta_i \right] + c|\varepsilon_l| - c|\varepsilon_{om}|$$

$$< 0 \quad (7)$$

As $|\varepsilon_l| \ll \Delta_j$, this equality is always false (compare to equation (5)).

If errors are relatively large, the sequence in which members want to stop the activity will depend mainly on the errors that they make. Thus, a democratic group is mainly influenced by the median error of all its members. As $|\varepsilon_i| \gg \Delta_j$, a group benefits from a leader decision, if:

$$c|\varepsilon_l| < c|\text{median}(\varepsilon_{om})| \Rightarrow 0.8\sigma_l < 0.8\frac{\sigma_{om}}{\sqrt{n}} \quad (8)$$

This inequality is true, if group size n is small and the median error of n inexperienced members is higher than the error of the one experienced leader.

Methods

Model of synchronization costs

We modelled the duration and resulting synchronization costs of a group activity session for despotic and democratic groups (Box 1 and Fig. 1). Optimal activity duration generally varies between members^{1-4,19,22}, so groups have to reach communal decisions about group activity duration¹⁻⁷. In a despotic group, group activity duration equals the optimal activity duration of the deciding despot, whereas a democratic group will stop the group activity when the number of members that want to stop reaches a majority. Individual members pay synchronization costs that increase with the difference between the actual group activity duration and the activity duration that would have been optimal for themselves^{1-4,22}.

Distribution of synchronization costs between members

In Box 2 we consider two extremes, namely: (1) all adult group members have similar time budgets⁴ (homogeneous group); or (2) all members differ so widely in their time budgets that there is no overlap in optimal stopping times between members^{4,22} (extremely heterogeneous group).

Groups with incomplete information

In Box 3, the costs to a group with an experienced leader are compared to those of a democratic group. We only consider homogeneous groups, because members mainly profit from an experienced leader if they are similar to it in their time budget requirements⁴. Two extreme cases are investigated: (1) errors are small; and (2) errors are large, relative to differences in optimal timing between members.

Received 30 August; accepted 1 November 2002; doi:10.1038/nature01294.

1. Prins, H. H. T. *Ecology and Behaviour of the African buffalo* (Chapman & Hall, London, 1996).
2. Cote, S. D., Schaefer, J. A. & Messier, F. Time budgets and synchrony in activity in muskoxen: the influence of sex, age and season. *Can. J. Zool.* **75**, 1628-1635 (1997).
3. Conradt, L. Could asynchrony in activity between the sexes cause inter-sexual social segregation in ruminants? *Proc. Roy. Soc. B* **265**, 1359-1363 (1998).

4. Ruckstuhl, K. E. To synchronise or not to synchronise: a dilemma for young bighorn males? *Behaviour* **136**, 805–818 (1999).
5. Stewart, K. J. & Harcourt, A. H. Gorillas vocalizations during rest periods—signals of impending departure. *Behaviour* **130**, 29–40 (1994).
6. Byrne, R. W. *On the Move* (eds Boinski, S. & Garber, P. A.) 491–518 (Univ. Chicago Press, Chicago, 2000).
7. Milton, K. *On the Move* (eds Boinski, S. & Garber, P. A.) 357–418 (Univ. Chicago Press, Chicago, 2000).
8. Poole, J. H., Payne, K., Langbauer, W. R. & Moss, C. J. The social contexts of some very low-frequency calls of African elephants. *Behav. Ecol. Sociobiol.* **22**, 385–392 (1988).
9. Black, J. M. Preflight signalling in swans—a mechanism for group cohesion and flock formation. *Ethology* **79**, 143–157 (1988).
10. Norton, G. W. *Primate Ecology and Conservation* (eds Else, J. G. & Lee, P. C.) 145–156 (Cambridge Univ. Press, Cambridge, 1986).
11. Stolba, A. *Entscheidungsfindung in Verbänden von Papio hamadryas*, 77–80. Thesis, Univ. Zurich (1979).
12. Boinski, S. & Campbell, A. F. Use of trill vocalizations to coordinate troop movement among white-faced capuchins—a 2nd field-test. *Behaviour* **132**, 875–901 (1995).
13. Stephens, D. W. & Krebs, J. R. *Foraging Theory* (Princeton Univ. Press, Princeton, New Jersey, 1986).
14. Krebs, J. R. & Kacelnik, A. *Behavioural Ecology: an Evolutionary Approach*, 3rd edn (eds Krebs, J. R. & Davies, N. B.) 105–136 (Blackwell, Oxford, 1991).
15. Cuthill, I. C. & Houston, A. I. *Behavioural Ecology: an Evolutionary Approach*, 4th edn (eds Krebs, J. R. & Davies, N. B.) 97–120 (Blackwell, Oxford, 1997).
16. Seeley, T. D. & Buhrman, S. C. Group decision making in swarms of honey bees. *Behav. Ecol. Sociobiol.* **45**, 19–31 (1999).
17. Conradt, L. & Roper, T. J. Activity synchronization and social segregation: a model. *Proc. Roy. Soc. B* **267**, 2213–2218 (2000).
18. Decastro, J. M. & Decastro, E. S. Spontaneous meal patterns of humans—influence of the presence of other people. *Am. J. Clin. Nutr.* **50**, 237–247 (1989).
19. Gompper, M. E. Sociality and asociality in white-nosed coatis (*Nasua narica*): foraging costs and benefits. *Behav. Ecol.* **7**, 254–263 (1996).
20. Grundgesetz fuer die Bundesrepublik Deutschland (BGB1. I S.1-1822, 1949).
21. Wagenknecht, E. *Der Rothirsch* (Ziemsen, Wittenberg Lutherstadt, Germany, 1980).
22. Clutton-Brock, T. H., Guinness, F. E. & Albon, S. D. *Red Deer: Behaviour and Ecology of two Sexes* (Univ. Chicago Press, Chicago, 1982).

Supplementary Information accompanies the paper on Nature's website
(<http://www.nature.com/nature>).

Acknowledgements We thank J. Soeding, D. Waxman and H. Kokko for comments on an earlier version of the model, and H. Prins and M. Manser for help with accessing references. L.C. was supported by a Royal Society University Research Fellowship.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to L.C.
(e-mail: l.conradt@sussex.ac.uk).

Larval stages of a living sea lily (stalked crinoid echinoderm)

Hiroaki Nakano^{*†}, Taku Hibino^{†‡}, Tatsuo Oji[§], Yuko Hara[‡] & Shonan Amemiya^{*‡}

^{*} Department of Integrated Biosciences, Graduate School of Frontier Sciences, University of Tokyo, Bldg. FSB-501, 5-1-5 Kashiwanoha, Kashiwa, Chiba, 277-8562, Japan

[‡] Department of Biological Sciences; and

[§] Department of Earth and Planetary Science; Graduate School of Science, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-0033, Japan

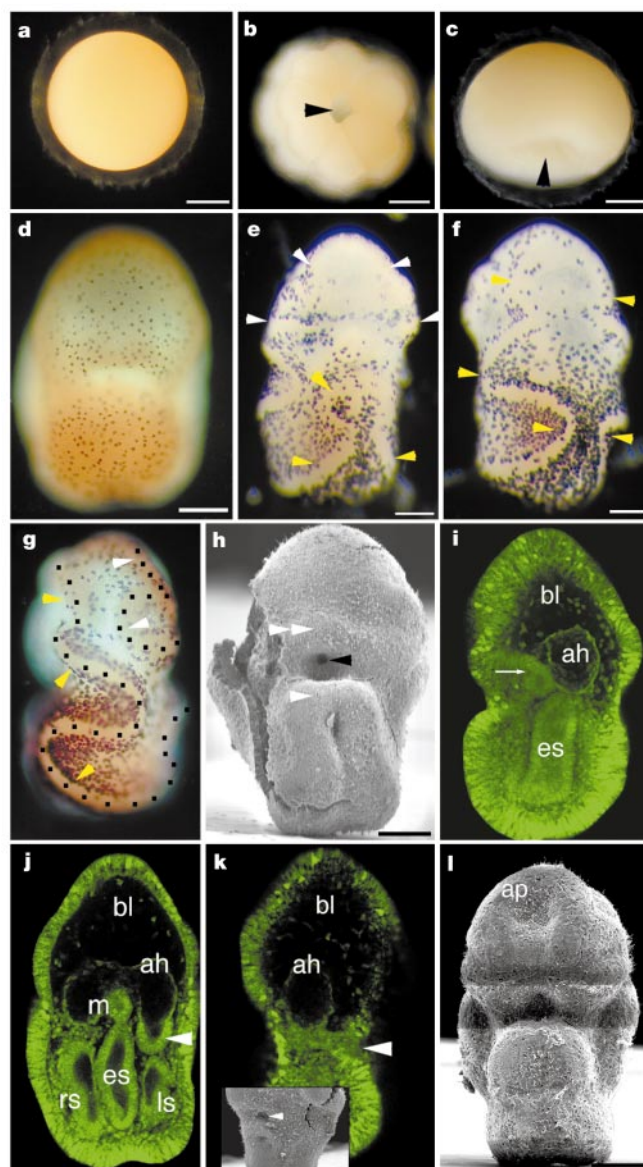
[†] These authors contributed equally to this work

The embryos and larvae of stalked crinoids, which are considered the most basal group of extant echinoderms^{1,2}, have not previously been described. In contrast, much is known about the development of the more accessible stalkless crinoids (feather stars)³, which are phylogenetically derived from stalked forms⁴. Here we describe the development of a sea lily from fertilization to larval settlement. There are two successive larval stages: the first is a non-feeding auricularia stage with partly longitudinal ciliary bands (similar to the auricularia and bipinnaria larvae of

holothurian and asteroid echinoderms, respectively); the second is a doliolaria larva with circumferential ciliary bands (similar to the earliest larval stage of stalkless crinoids). We suggest that a dipleurula-type larva is primitive for echinoderms and is the starting point for the evolution of additional larval forms within the phylum. From a wider evolutionary viewpoint, the demonstration that the most basal kind of echinoderm larva is a dipleurula is consistent with Garstang's auricularia theory⁵ for the phylogenetic origin of the chordate neural tube.

Ripe males and females of the sea lily *Metacrinus rotundus* were collected by fishnet from a depth of 100–150 m in Suruga Bay, Japan, on 22 September 1998 (specimens 1) and 20 September 2000 (specimens 2), and in Sagami Bay, Japan, on 27 September 2002 (specimens 3). The crinoids were transferred to a laboratory aquarium (at 14 °C), where the females spawned at about 19:00 on the date of collection. The eggs were fertilized by sperm obtained by dissecting the testes of a ripe male. The fertilized eggs developed into cleavage stage embryos, but the embryos from specimens 1 and 3 ceased development at this stage. Embryos derived from specimens 2 developed into larval stages.

The unfertilized eggs, which have a diameter of $350 \pm 9 \mu\text{m}$ (mean $\pm$ s.d., $n = 30$) are pale yellow, opaque, and slightly lighter than seawater. The fertilized eggs elevate a conspicuous rough



fertilization envelope (Fig. 1a). At 12 °C the first cleavage takes place 2.5 h after fertilization, and the next few cleavages occur at intervals of about 1 h. Cleavage is holoblastic, radial and equal. By the 32-cell stage, a large pore forms (Fig. 1b), possibly equivalent to the vegetal pore found at the vegetal pole of feather star embryos³. By 24 h, a gastrula results from invagination at the vegetal pole (Fig. 1c).

During the next few hours the blastopore closes, while the embryo becomes uniformly ciliated and begins to rotate inside the fertilization envelope. The developmental stage that hatches (starting at about 38 h) has a uniformly ciliated exterior. After hatching (Fig. 1d), the posterior half of the body turns more orange than the pale yellow anterior half, scattered pigment spots appear, and slight ridges prefigure the later development of the ciliary bands.

By 3.5 days, the larva, which can now be termed an auricularia, has an anterior and a posterior ciliated band (Fig. 1e–g), each characterized by an absence of pigment spots. It is possible that the two bands are contiguous, but where they approach most closely at the anterior end, the possible connecting regions are vague owing to the shortness and sparseness of the cilia there. The overall shape of the larva and the course of the ciliary bands at this stage resemble those of a sea-cucumber auricularia larva and a starfish bipinnaria larva. The ventral side of the 3.5-day larva (Fig. 1h) is indented by a vestibular invagination, in the roof of which is a mouth invagination that is not connected with the rest of the gut (Fig. 1i). Thus the sea lily auricularia does not feed (nor do the later developmental stages described here). The vestibular invagination is bounded anteriorly by an oral hood (bearing the adoral ciliary band) and posteriorly by an anal hood, although no signs of an anus are visible at this stage (Fig. 1h and i).

Internally, part of the blastocoel space is conspicuous at the anterior end of the 3.5-day embryo (Fig. 1i–k). At 3.5 days, the archenteron is divided into the enteric sac and three coelomic

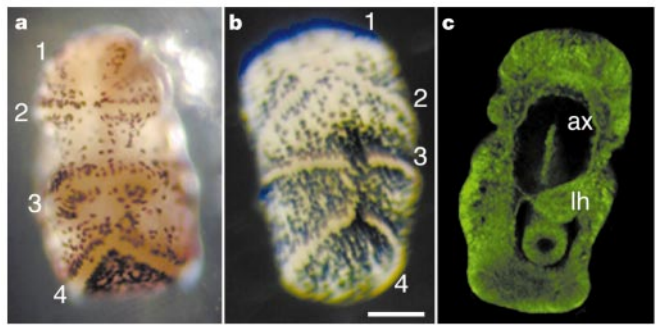


Figure 2 Doliolaria larvae of *M. rotundus*. Anterior is towards the top. Scale bar, 100 μ m. **a, b**, Ventral and dorsal views, respectively, of 10-day doliolaria larvae showing four circumferential ciliary bands (numbered 1–4). **c**, CLSM showing a frontal view of a 10-day doliolaria larva. The left hydrocoel (lh) has separated from the axocoel (ax), shifted to a midventral position, and become horseshoe-shaped.

pouches (an axo-hydrocoel and a left and right somatocoel) (Fig. 1j). A thickening of the posterior wall of the axo-hydrocoel marks the beginning of the left hydrocoel (Fig. 1j). The left side of the axo-hydrocoel has established communication with the exterior via the hydropore at the dorsal left side (Fig. 1k).

Between 3.3 and 6 days (Fig. 1l), the adhesive pit appears as a depression at the anteroventral side, the mouth invagination closes, and the larvae later begin pre-settlement swimming behaviour with their adhesive pit close to the substrate.

Between 6 and 10 days, the overall dimensions of the larvae shrink, and the ciliary bands become rearranged (Fig. 2a and b). Some parts of the bands break up, whereas others fuse together, eventually forming four circumferential ciliary bands. The 10-day-old larva can thus be defined as a doliolaria of the sort characteristic of feather stars, brittle stars and sea cucumbers^{3,6,7}.

Internally (Fig. 2c), the left hydrocoel had separated from the axocoel, shifted to a midventral position, and assumed a horseshoe shape during the transition into the doliolaria larva; moreover, enlargement of the coeloms had largely obliterated the blastocoel space. The doliolaria larvae settled by attachment to the substrate, starting at 10 days and for several days thereafter. The settled larvae lost their circumferential ciliary bands and assumed an oval shape. At this point development ceased and the settled larvae died.

Before the present study, known dipleurula-type larvae included the tornaria larvae of enteropneust hemichordates⁸ as well as the earliest larval stages (auricularia, bipinnaria and pluteus larvae) of all echinoderm classes except for crinoids⁹. Although these similarities were widely accepted as key evidence for a close evolutionary relationship between hemichordates and echinoderms^{10,11}, the apparent lack of a dipleurula-type larva in crinoids (the most basal class of extant echinoderms) was paradoxical. Our present work removes this paradox and leads us to propose that: (1) within the deuterostomes, a dipleurula larva is a synapomorphy uniting hemichordates and echinoderms, (2) a dipleurula followed by a doliolaria is a synapomorphy for echinoderms, (3) a secondary loss of the dipleurula has occurred in stalkless crinoids (though not completely in some species¹²), and (4) a secondary loss of the doliolaria has taken place in starfish and sea urchins. Our work is also consistent with Garstang's proposition^{5,13} (his auricularia theory) that the chordate central nervous system is derived from the ciliary bands, probably with basal neurociliary innervation, of an ancestral deuterostome with a dipleurula-like body plan.

Received 2 September; accepted 14 October 2002; doi:10.1038/nature01236.

1. Paul, C. R. C. & Smith, A. B. The early radiation and phylogeny of echinoderms. *Biol. Rev.* **59**, 443–481 (1984).
2. Janies, D. Phylogenetic relationships of extant echinoderm classes. *Can. J. Zool.* **79**, 1232–1250 (2001).

Figure 1 Early development through auricularia larvae of the sea lily *Metacrinus rotundus* cultured in the dark in artificial seawater at 12 °C (Jamarin, Osaka) containing 50 μ g ml⁻¹ streptomycin and 100 U ml⁻¹ penicillin. The culture water was changed every few days. For all larval stages, anterior is toward the top. Scale bars, 100 μ m. **a**, Fertilized egg within a rough fertilization envelope. **b**, 32-cell embryo with probable vegetal pore (arrowhead). **c**, Approximate side view of mid-gastrula (27.5 h after fertilization) showing the blastopore (arrowhead). **d**, Ventral view of swimming 3-day larva. **e**, Ventral view of 4-day auricularia larva with anterior (white arrowheads) and posterior (yellow arrowheads) ciliary bands. **f**, Dorsal view of previous larva with yellow arrowheads indicating posterior ciliary band. **g**, Lateral view of 4-day auricularia larva (different larva from **e** and **f**) with anterior (white arrowheads) and posterior (yellow arrowheads) ciliary bands indicated with black dots. **h**, Scanning electron micrograph (SEM) of 3.5-day larva that had been fixed for 1 h in 2.5% glutaraldehyde in seawater, postfixed for 1 h in 1% osmium in seawater, dehydrated in ethanol, transferred to butanol, freeze-dried and coated with gold (the right side was damaged during this procedure). This ventral view shows a mouth invagination (black arrowhead) within the vestibular invagination bordered by the oral and anal hoods (white twin and single arrowheads, respectively). **i**, Confocal laser scanning micrograph (CLSM) showing a sagittal view of 3.5-day larvae fixed overnight at 4 °C in 4% paraformaldehyde, 0.1 M MOPS at pH 7.5 and 0.5 M NaCl before transfer via ethanol to 50:50 benzyl benzoate/benzyl alcohol. Note that the mouth invagination (arrow) is not connected to the enteric sac (es). Other abbreviations: ah, axo-hydrocoel; bl, blastocoel space. **j**, CLSM showing a frontal view of the larva in **i**. The archenteron divides into the enteric sac and three coelomic pouches, presumably the axo-hydrocoel, the left somatocoel (ls) and the right somatocoel (rs). The left posterior portion of the axo-hydrocoel evaginates and the wall thickens, defining the future left hydrocoel (arrowhead). **m**, invagination of the mouth. **k**, CLSM showing a left lateral sagittal view of the larva in **j**. The left side of the axo-hydrocoel has established communication with the exterior via the hydropore at the dorsal left side (arrowhead). inset: SEM of the posterodorsal side showing the hydropore (arrowhead). **l**, SEM showing the ventral view of 4-day auricularia larva. The adhesive pit (ap) appears as a depression at the anteroventral side. No opening of the mouth or the anus is present.

- Holland, N. D. *Reproduction of Marine Invertebrates* (Echinoderms and Lophophorates) (eds Giese, A. C., Pearse, J. S. & Pearse, V. B.) Vol. 4, 247–299 (Boxwood Press, Pacific Grove, California, 1991).
- Foote, M. Morphological diversity in the evolutionary radiation of Paleozoic and post-Paleozoic crinoids. *Paleobiology* 25 suppl. S SPR, 1–115 (1999).
- Garstang, W. Preliminary note on a new theory of the phylogeny of the Chordata. *Zool. Anz.* 17, 122–125 (1894).
- Hendler, G. *Reproduction of Marine Invertebrates* (Echinoderms and Lophophorates) (eds Giese, A. C., Pearse, J. S. & Pearse, V. B.) Vol. 4, 355–511 (Boxwood Press, Pacific Grove, California, 1991).
- Smiley, S., McEuen, F. S., Chaffee, C. & Krishnan, S. *Reproduction of Marine Invertebrates* (Echinoderms and Lophophorates) (eds Giese, A. C., Pearse, J. S. & Pearse, V. B.) Vol. 4, 663–750 (Boxwood Press, Pacific Grove, California, 1991).
- Nielsen, C. *Animal Evolution* (Oxford Univ. Press, Oxford, 1995).
- Strathmann, R. R. Introduction to function and adaptation in echinoderm larvae. *Thalass. Jugoslav.* 10, 321–339 (1974).
- Metschnikoff, E. Über die systematische Stellung von Balanoglossus. *Zool. Anz.* 4, 139–143 (1881).
- Hyman, L. H. *The Invertebrates* (Echinodermata: The coelomate Bilateria) Vol. 4 (McGraw-Hill, New York, 1955).
- Lacalli, T. C. & West, J. E. Ciliary band formation in the doliolaria larva of *Florumetra 1* The development of normal epithelial pattern. *J. Embryol. Exp. Morph.* 96, 303–323 (1986).
- Gee, H. *Before the Backbone* (Chapman & Hall, London, 1996).

Acknowledgements We thank C. Kitazawa, H. Nakagawa and A. Hirata for technical assistance with scanning electron microscopy, A. Nishino for discussions, and N. D. Holland for reading the manuscript and making valuable comments. This work was supported by grants from The Japan Science Society to Y.H., and from the Sumitomo Foundation and MEXT Japan to S.A.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.A. (e-mail: shonan@k.u-tokyo.ac.jp).

Robust judgement of inter-object distance by an arthropod

Jan M. Hemmi & Jochen Zeil

Centre for Visual Sciences, Research School of Biological Sciences, Australian National University, PO Box 475, Canberra, Australian Capital Territory 2601, Australia

Animals use several strategies for depth vision, reflecting the constraints imposed by body size, the structure of the visual system and the visual geometry of the environment¹. Arthropods in particular have restricted depth perception, because they are small and possess closely set, low-resolution compound eyes. Yet, here we show that fiddler crabs defending their burrows from conspecifics can judge how close other crabs are to their burrow. When confronted with small dummy crabs, the burrow owners assess the dummy's position and motion relative to their burrow and not relative to themselves—in other words, by using an allocentric rather than an egocentric frame of reference. Irrespective of their own distance from the dummy, the likelihood that the crabs rush back to defend their burrow increases strongly as the dummy approaches the burrow. In addition, the mean dummy–burrow distance at which the crabs respond is constant and independent of the dummy's direction of approach. We propose that to solve this sophisticated task of relative distance judgement, the crabs combine visual information on dummy position and direction with information on burrow location acquired during path integration². In doing so, the crabs, like humans³, make clever use of the visual geometry of their environment.

The fiddler crab *Uca vomeris* lives in dense colonies on tropical and subtropical mud and sandflats, commonly with between 10 and 20 animals per square metre. Each individual operates from its own burrow, which it defends vigorously against other crabs^{4–7}. Burrow

defence is not a trivial task because in most cases foraging crabs cannot see the entrance to their burrow owing to perspective foreshortening⁸. To determine how fiddler crabs detect and assess conspecifics approaching their burrows, we moved small dummy crabs across the ground towards a burrow while the burrow owner was away feeding. We subsequently reconstructed from video tapes the exact spatial constellation of the dummy, the crab and its burrow in a frame-by-frame analysis to determine the timing of burrow-defence responses and the visual cues triggering them.

When confronted with a dummy crab approaching their burrow, burrow owners respond by dashing back to their burrows, just as they would in response to real crab intruders. Figure 1 documents two such burrow-defence responses. In the first scene (Fig. 1a), the crab slowly retreats to its burrow once the dummy has approached to within 21 cm of the entrance. The crab then remains close to the burrow until the dummy has left. In the second scene (Fig. 1b), the same crab is on the other side of the burrow when the dummy approaches and responds when the dummy has reached a similar dummy–burrow distance (26 cm), even though the crab–dummy distance at this point is still much larger than in the first scene (Fig. 1a). The crab responds with a fast home run that takes it all the way back to the burrow. After the dummy has reached its closest approach, the crab begins to move away from the burrow. But as soon as the dummy returns from the end of its track, the crab rushes home for the second time, where it remains until the dummy has moved out of the area.

We find that the response probability increases strongly as the dummy approaches the burrow (Fig. 2, black line), resulting in a cumulative response probability that reaches 66% at 10 cm (Fig. 2, grey line). As there is no defined dummy–burrow distance for trials in which the crabs did not react, we used the variable 'track distance', which measures the closest distance between the dummy track and the crab's burrow (see Fig. 3a) and varied between 5 and 55 cm, to test for the effect of the dummy–burrow distance on the crab's response probability. Statistical analysis shows that the track distance is the main factor that determines whether a crab responds to these dummies or not ($P < 0.001$). The probability of response decreases from over 80% for track distances of 5–10 cm, to below 20% for track distances of more than 50 cm.

The crabs tend to stay close to their burrows and the track distance is therefore correlated with the distance between the crab and the dummy track. To ensure that response probability is indeed determined by the dummy–burrow distance and not by the crab–dummy distance, we removed the (burrow) track distance from the

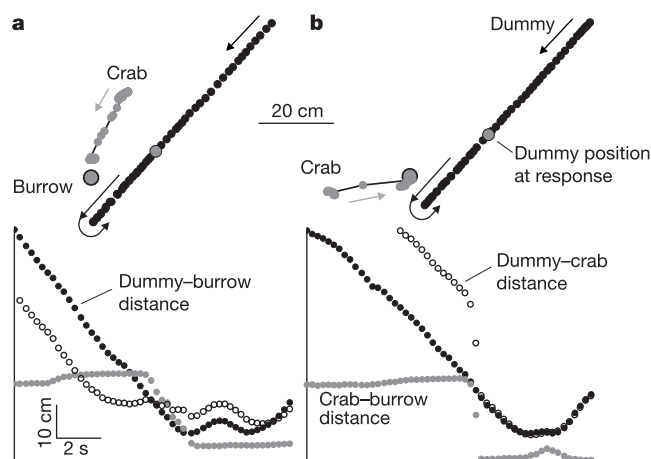


Figure 1 Two examples of a fiddler crab (*Uca vomeris*) responding with burrow defence to an approaching crab dummy. The top section of each panel shows the dummy, the crab and the burrow from above. Dots mark the dummy and crab positions at 240-ms intervals.

model and fitted the crab–track distance instead. Again, we could not use crab–dummy distance directly because it is not defined for trials in which crabs did not react. The crab–track distance has no explanatory power (generalized linear mixed model (GLMM): degrees of freedom (d.f.) = 1, Wald/d.f. = 0.37, $P = 0.543$), even when we allow for a difference in response depending on whether the dummy approaches from the crab's side of the burrow or not (GLMM: d.f. = 1, Wald/d.f. = 0.29, $P = 0.748$). The effects of additional variables in the statistical model are summarized in Methods. The distance between the dummy and the burrow is thus the most important predictor of whether a crab reacts or not.

We examined what determines the timing of the response by investigating the parameters that affect the response distance, which is the dummy–burrow distance at the moment the crabs decide to respond. Statistical analysis showed that the average response distance was 23.8 cm and was unexpectedly constant across many factors (see Methods and Fig. 3a). Most notably, the crabs react at almost equal dummy–burrow distances, irrespective of the dummy's approach direction and consequently of the distance between crab and dummy (Fig. 3b). The statistical model identifies a small influence of the dummy's direction of approach (track angle), which just reaches significance ($P = 0.028$). But this effect is very small and is caused by the crabs reacting slightly late whenever the dummy approaches from beyond the burrow (Fig. 3b).

Whereas the crabs keep the response distance constant, irrespective of the dummy's approach direction, they are able to adjust the response distance (dummy–burrow distance) according to their own distance from the burrow. For small track distances, there is a positive effect of the crab–burrow distance on the response distance. The further away the crabs are from the burrow, the earlier they respond to the dummy ($P < 0.001$). The crabs thus know their own distance from the burrow, most probably through path integration². The main conclusion from these results is that fiddler crabs can measure the distance between a dummy crab and their burrow, irrespective of the relative angle under which they view the dummy and the burrow.

Are the crabs also sensitive to the movement direction of the dummies? To investigate this possibility, we selected all experiments in which the dummy moved at least 5 cm past the closest point to the burrow before changing direction at the return point, including experiments where the crabs did not react at all or reacted late after the dummy had moved past the closest point (Fig. 4, inset). If the crabs discriminate between approaching and retreating dummies, then we expect the response probability to be higher for segments 1 and 3 than for segments 2 and 4.

The data in Fig. 4 clearly show that the crabs distinguish between approaching and retreating dummies. The response probability is significantly higher for track segments 1 and 3 than for segments 2 and 4 (GLMM: d.f. = 3, Wald/d.f. = 11.95, $P < 0.001$). On the basis of the pair-wise standard errors of the linear transformation of the model, all four segments are significantly different from each

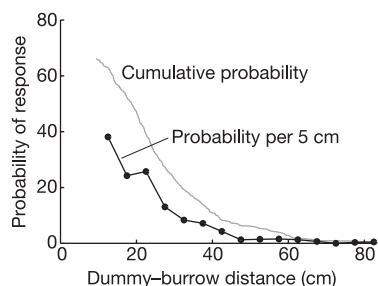


Figure 2 The crabs' response probability (%) depends on the dummy–burrow distance. The black unbroken line shows the probability that the crabs react in a 5-cm interval. The grey line shows the corresponding cumulative response probability.

other. Because the dummies always moved through the segments in the order 1 to 4, one might expect the response to decline towards later segments, because the more sensitive crabs might react earlier in the sequence. The most important comparison is therefore between segments 2 and 3, which clearly confirms the hypothesis that the crabs are sensitive to the dummy's direction of motion. In addition, the response probability is higher in segment 3 than in segment 1, which suggests that the crabs are sensitized by the change of the dummy's direction at the return point.

We suggest that the onset of burrow defence in fiddler crabs depends on two variables: the distance of an intruder to the owner's burrow; and the intruder's direction of motion relative to the burrow—that is, whether the intruder approaches it or not. Fiddler crabs are thus able to assess the distance and the motion direction of other crabs not only in an egocentric frame of reference, but also relative to a distant, invisible location in the world, in other words, a geocentric frame of reference.

The predictable visual geometry of the flat world fiddler crabs inhabit allows them to use a simple solution for what seems to be the complex geometrical problem of judging the distance between two objects independently of the viewing angle and distance. Theoretically, in the crabs' visual environment the distance of objects on the ground can be determined by retinal position (elevation) alone^{1,3,9}. The crabs carry their eyes at a constant height above ground on long, vertical eye stalks, and when foraging away from the burrow they

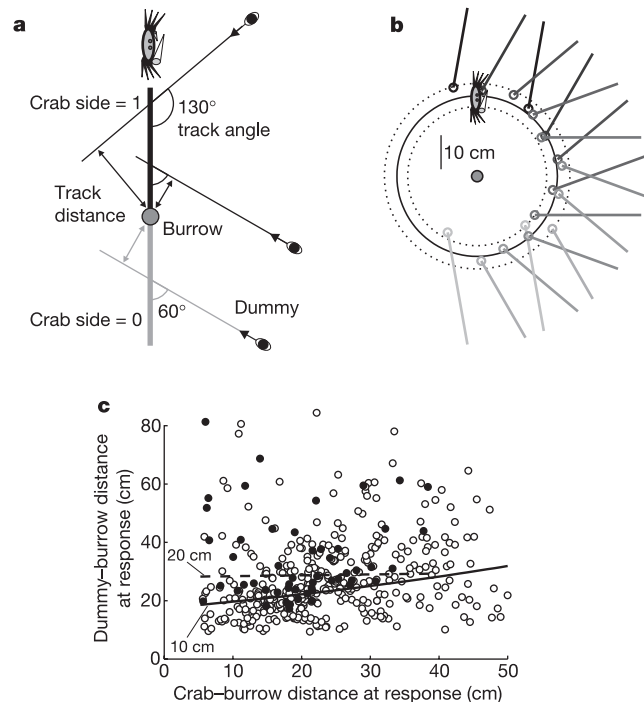


Figure 3 The timing of the burrow defence response. **a**, Definition of parameters. The approach direction (track angle) was fitted as a factor with nine directions (20° bins). **b**, The dummy–burrow distance at the time of the response is constant. The solid lines show the dummy's trajectory and the large dots mark its position at the time of the response as fitted by the statistical model. The response distance is shown for each of the nine approach directions (track angle) and the two values of crab side. Increasing track angles are shown in progressively lighter greys. The unbroken and dotted circles represent the mean reaction distance and ± 2 s.e.m., respectively. Fitted values were calculated at the mean of all other parameters in the statistical model (crab–burrow distance = 23.4 cm). **c**, For small track angles, the crabs adjust the response distance according to their own distance from the burrow (open circles, track distance < 15 cm; filled circles, track distance ≥ 15 cm). The corresponding model fit is shown as an unbroken line for a track distance of 10 cm and as a dashed line for a track distance of 20 cm.

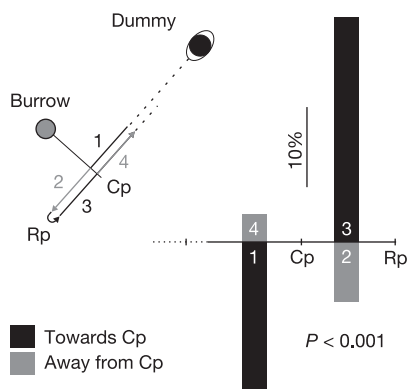


Figure 4 The probability of response for four length-matched segments around the closest point (Cp) of the dummy track to the crab's burrow (see inset). Only the first response of a crab was counted. If a crab responded in a given segment, then its response during the remainder of the sequence was ignored. Experiments in which the crabs reacted before the dummy reached segment 1 were not included. The number of trials contributing to segments were 232, 176, 162 and 123, respectively. Rp, the return point of the dummy at the end of the track.

keep their longitudinal body axis aligned with the home vector². The burrow environment is thus always viewed by the lateral retina. For a given crab–burrow distance, the crabs' ability to measure the distance between an object and the burrow could thus be based on a single, wide-field, motion-sensitive neuron (distance neuron) with a gradient of sensitivity to small objects that increases towards the retinal position of the burrow (Fig. 5a). Owing to perspective distortion, this gradient changes with crab–burrow distance so that foraging crabs would need to use a few such cells, each tuned to a certain range of crab–burrow distances (Fig. 5b). We suggest that the crabs use the state of their path integrator² to select which of these neurons (or which combination) is most appropriate for the current crab–burrow distance.

For the crab's second task, namely to decide whether the dummy is approaching the burrow or not, there are two alternative solutions. The simplest solution would be to monitor the output of the distance neurons, because objects that approach the burrow would lead to an increase in their response. Alternatively, the directional selectivity could be achieved by appropriately aligned local motion detectors serving as input elements to the distance neurons, in a

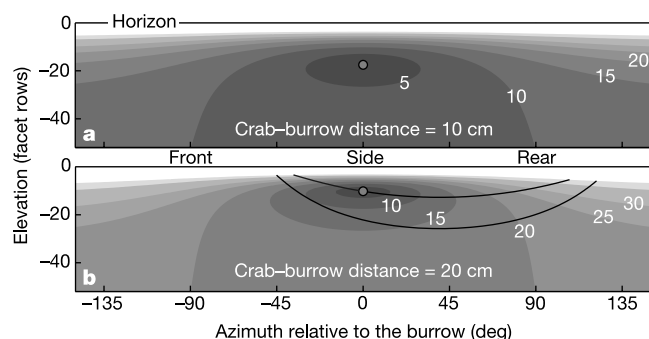


Figure 5 The effect of the crab–burrow distance on the mapping of concentric circles around the burrow onto the ommatidial array of a fiddler crab eye^{13,14}. The x axis shows azimuth in degrees, the y axis shows elevation in units of horizontal facet rows. The scaling is such that individual ommatidia take up equal space along both axes. Contours have been drawn for concentric circles with a radius of 5, 10, 15, 20, 25, 30, 40 and 50 cm. The distance map is shown for crab–burrow distances of 10 cm (a) and 20 cm (b). Two example dummy paths (unbroken lines), with a track angle of 50° and a track distance of 0 and 10 cm respectively, are also shown in b.

similar way to neurons involved in optic flow processing¹⁰. In the predictable visual environment of fiddler crabs, the difficult problem of measuring the distance between two distant objects, irrespective of the viewing angle, can thus be solved by a relatively simple matched neural filter¹¹. □

Methods

Apparatus

Crab dummies consisted of small black or white plastic cylinders with a width of 1.5 or 2.25 cm and a height of 0.8, 1.2, 1.6 or 2 cm. We carried out 484 of the total 633 experiments with a dummy size of 2.25 × 1.2 cm (width × height). The dummies were mounted on a thin, transparent perspex sled with the same width as the dummies, but twice the length to allow smooth motion on the ground. The dummies were moved along a straight line, the 'dummy track', by manually pulling on a monofilament line, which was fed around two tent pegs. The dummy movements could thus be controlled by an observer sitting about 5–6 m away. The dummy speed was roughly constant for a given experiment and ranged from 3 to 14 cm s⁻¹ between experiments (mean ± s.d. 7.8 ± 1.9 cm s⁻¹). We moved the dummy 10 cm past a focal crab's burrow but also monitored the reactions of neighbouring crabs by filming an area of about 1 m² around the focal crab's burrow from above. The crabs very quickly and fully habituate to the presence of both the camera and the observer.

Image analysis

On the basis of the x and y coordinates of the crabs, the dummy and the burrow, a response was considered to have started in a given frame if a crab moved at least 0.66 cm towards its burrow in the 240-ms time interval preceding this frame and at least 2 cm over a three-frame interval (720 ms) starting at the previous frame. We only analysed crab responses when the resident crab was at least 5 cm away from its burrow at the time of reaction. Our final analysis includes 633 dummy presentations from 25 animals. The crabs reacted in 493 of these experiments; however, responses were only counted if the crab reacted before the dummy had reached its closest point to the burrow for the first time (419 of the 493 responses or 85%). This was necessary because of the 74 late reactions, most of which (60) happened after the dummy turned around at the end of the track to move back to its starting position. The analysis in Fig. 4 is based on these late reactions.

Statistics

All of the statistical models used in this analysis allowed us to account for the repeated dummy presentations per crab (crab-to-crab variation). The significance of individual terms in all models was tested by calculating the Wald statistic associated with dropping the term from the model. To test for factors that influence the crabs' probability of response, we used the GLMM¹² in Genstat v4.2 (VSN International Ltd). No interactions reached significance and the final model had the following form ($n = 633$, $\text{Logit}(p) = \log(p/(1 - p))$): $\text{Logit}(p) \approx \beta_0 + \beta_1(\text{track distance}) + \beta_2(\text{track angle}) + \beta_3(\text{crab side}) + \beta_4(\text{presentation repeat}) + \text{error}$. The effect of track distance (d.f. = 1, Wald/d.f. = 15.64, $P < 0.001$) is fully discussed in the text. Decreasing track angles (see Fig. 3a) led to a decrease in the reaction probability from 70–80% for angles larger than 100° to about 40% for dummies that approach from beyond the burrow (d.f. = 8, Wald/d.f. = 4.89, $P = 0.001$). The response probability was about 10% higher for experiments in which crab side (see Fig. 3a) had a value of 1 (d.f. = 1, Wald/d.f. = 5.71, $P = 0.017$). Response probability decreased over 50 presentations from over 80% to about 40% (presentation repeat, d.f. = 5, Wald/d.f. = 5.47, $P < 0.001$).

To examine the response timing, we carried out a restricted maximum likelihood analysis (REML) using Genstat v4.2 that resulted in the following model ($n = 419$; Fig. 3): $\log(\text{dummy–burrow distance}) \approx \beta_0 + \beta_1(\text{track distance}) + \beta_2(\text{crab–burrow distance}) + \beta_3(\text{track distance} \times \text{crab–burrow distance}) + \beta_4(\text{presentation repeat}) + \beta_5(\text{track angle}) + \beta_6(\text{crab side}) + \text{error}$. The effects of track angle (d.f. = 8, Wald/d.f. = 2.15, $P = 0.028$) and the interaction between track distance and crab–burrow distance (d.f. = 1, Wald/d.f. = 12.67, $P < 0.001$) are discussed in the text. The effect of crab side (d.f. = 1, Wald/d.f. = 3.76, $P = 0.053$) was a small, 2.6-cm increase in the response distance (see Fig. 3b). Response distance decreased with presentation repeat from 31 cm to 21 cm over the course of 50 presentations (d.f. = 5, Wald/d.f. = 7.16, $P < 0.001$).

Neither the dummy brightness (black or white) nor the dummy speed had any influence on the response probability (GLMM: Wald/d.f. = 0.18, $P = 0.67$; Wald/d.f. = 2.75, $P = 0.10$) or on the response timing (REML: Wald/d.f. = 0.59, $P = 0.44$; Wald/d.f. = 0.31, $P = 0.58$).

Received 16 August; accepted 7 October 2002; doi:10.1038/nature01247.

- Collett, T. S. & Harkness, L. I. K. *Analysis of Visual behaviour* (eds Ingle, D. J., Goodale, M. A. & Mansfield, R. J. W.) 111–176 (MIT Press, Cambridge, MA, 1982).
- Zeil, J. Homing in fiddler crabs (*Uca lactea annulipes* and *Uca vomeris*, Ocypodidae). *J. Comp. Physiol. A* **183**, 367–377 (1998).
- Ooi, T. L., Wu, B. & He, Z. J. Distance determined by the angular declination below the horizon. *Nature* **414**, 197–200 (2001).
- Crane, J. *Fiddler Crabs of the World (Ocypodidae: genus Uca)* (Princeton Univ. Press, Princeton, NJ, 1975).
- Hyatt, G. W. & Salmon, M. Combat in the fiddler crabs *Uca pugnator* and *U. pugnax*: a quantitative analysis. *Behaviour* **65**, 182–211 (1978).
- Jennions, M. D. & Backwell, P. R. Y. Residency and size affect fight duration and outcome in the fiddler crab *Uca annulipes*. *Biol. J. Linn. Soc.* **57**, 293–306 (1996).
- Montague, C. L. A natural history of temperate western Atlantic fiddler crabs (Genus *Uca*) with

reference to their impact on the salt marsh. *Contrib. Mar. Sci.* **23**, 25–55 (1980).

8. Zeil, J. & Layne, J. *Crustacean Experimental Systems in Neurobiology* (ed. Wiese, K.) 227–247 (Springer, Heidelberg, 2002).
9. Zeil, J., Nalbach, G. & Nalbach, H.-O. Eyes, eye stalks, and the visual world of semi-terrestrial crabs. *J. Comp. Physiol. A* **159**, 801–811 (1986).
10. Krapp, H. G., Hengstenberg, B. & Hengstenberg, R. Dendritic structure and receptive-field organization of optic flow processing interneurons in the fly. *J. Neurophysiol.* **79**, 1902–1917 (1998).
11. Wehner, R. 'Matched filters'—neural models of the external world. *J. Comp. Physiol. A* **161**, 511–531 (1987).
12. Schall, R. Estimation in generalized linear models with random effects. *Biometrika* **78**, 719–727 (1991).
13. Zeil, J. & Al-Mutairi, M. M. The variation of resolution and of ommatidial dimensions in the eyes of the fiddler crab *Uca lactea annulipes* (Ocypodidae, Brachyura, Decapoda). *J. Exp. Biol.* **199**, 1569–1577 (1996).
14. Land, M. & Layne, J. The visual control of behaviour in fiddler crabs I. Resolution, thresholds and the role of the horizon. *J. Comp. Physiol. A* **177**, 81–90 (1995).

Acknowledgements We thank P. Dixon, L. Trott and L. Howlett for help; J. Wood for advice; and M. R. Ibbotson and M. V. Srinivasan for comments on the manuscript. The work was supported by a postdoctoral fellowship from the Swiss National Foundation to J.M.H. and in part by an Human Frontier Science Program (HFSP) grant.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.M.H. (e-mail: jan.hemmi@anu.edu.au).

An active DNA transposon family in rice

Ning Jiang^{*†}, Zhirong Bao^{†‡}, Xiaoyu Zhang^{*}, Hirohiko Hirochika[§], Sean R. Eddy[‡], Susan R. McCouch^{||} & Susan R. Wessler^{*}

^{*} Department of Plant Biology, University of Georgia, Athens, Georgia 30602, USA

[‡] Howard Hughes Medical Institute and Department of Genetics, Washington University School of Medicine, St Louis, Missouri 63110, USA

[§] National Institute of Agrobiological Resources, Tsukuba, Ibaraki 305, Japan

^{||} Department of Plant Breeding, Cornell University, Ithaca, New York 14853, USA

[†] These authors contributed equally to this work

The publication of draft sequences for the two subspecies of *Oryza sativa* (rice), *japonica* (cv. Nipponbare) and *indica* (cv. 93-11)^{1,2}, provides a unique opportunity to study the dynamics of transposable elements in this important crop plant. Here we report the use of these sequences in a computational approach to identify the first active DNA transposons from rice and the first active miniature inverted-repeat transposable element (MITE) from any organism. A sequence classified as a *Tourist*-like MITE of 430 base pairs, called *miniature Ping* (*mPing*), was present in about 70 copies in Nipponbare and in about 14 copies in 93-11. These *mPing* elements, which are all nearly identical, transpose actively in an *indica* cell-culture line. Database searches identified a family of related transposase-encoding elements (called *Pong*), which also transpose actively in the same cells. Virtually all new insertions of *mPing* and *Pong* elements were into low-copy regions of the rice genome. Since the domestication of rice *mPing* MITEs have been amplified preferentially in cultivars adapted to environmental extremes—a situation that is reminiscent of the genomic shock theory for transposon activation³.

Rice is the most important crop for human nutrition in the world. At 430 megabase pairs (Mb), it also has the smallest genome among the agriculturally important cereals (including maize, sorghum, barley and wheat)⁴. For these reasons, rice is the focus of several genome-sequencing projects^{1,2,5}. Computer-assisted analyses of rice genomic sequence indicate that, despite its small size, over 40% is repetitive DNA and most of this is related to transposable elements^{1,2}. The class 1 long-terminal repeat (LTR) retro-

transposons comprise the largest component of transposable elements in the rice genome (14% of the genomic DNA) but, numerically, MITEs form the largest group with over 100,000 elements divided into hundreds of families comprising about 6% of the genome^{6,7}. MITEs are the predominant transposable element associated with the non-coding regions of the genes of flowering plants, especially grasses, and have been found in several animal genomes including *Caenorhabditis elegans*, mosquitoes, fish and human (reviewed in ref. 8).

Structurally, MITEs are reminiscent of non-autonomous DNA (class 2) elements with their small size (<600 base pairs) and short (10–30 bp) terminal inverted repeats (TIRs). But their high copy number (up to 10,000 copies per family) and target-site preference (for TA or TAA) distinguish them from most previously described non-autonomous DNA elements⁸. Non-autonomous elements, which make up a significant fraction of eukaryotic genomes, have been classified into families according to the transposase responsible for their mobility. But classifying MITEs in this way is problematic because no actively transposing MITE had been reported in any organism. Instead, the tens of thousands of plant MITEs have been classified into two superfamilies on the basis of the similarity of their TIRs and their target site duplication (TSD): *Tourist*-like MITEs and *Stowaway*-like MITEs^{7,9,10}. Much evidence links *Tourist* and *Stowaway* MITEs with two superfamilies of transposases, *PIF/Harbinger* and *Tc1/mariner*, respectively^{9–11}.

Because newly transposed elements should be identical, we considered that an active MITE family would be characterized by extremely low intrafamily sequence divergence. The availability of almost half of the Nipponbare genome (187 Mb by 24 December 2001) in public databases (<http://rgp.dna.affrc.go.jp>) provided the possibility of searching for repeat families with the structural features of MITEs and with very low intrafamily sequence divergence. We first identified 1,257 repeat families with RECON¹², a program for the *de novo* identification of repeat families. Manual inspection of these sequences subsequently indicated that a repeat of 430 bp named *mPing* was a candidate for an active MITE. The TSDs (the trinucleotide TAA or TTA) and TIRs of *mPing* indicated that it is a *Tourist*-like MITE (Fig. 1). Of the 36 independent copies mined from 270 Mb of Nipponbare sequence, 26 were identical. By contrast, only eight complete copies of *mPing* were found in the 361 Mb of contig sequence of the *indica* cultivar 93-11 (ref. 2). From the frequency of elements recovered per megabase, we estimated the whole genome of Nipponbare and 93-11 to contain 70 and 14 copies of *mPing*, respectively (Methods).

The only rice transposable elements shown previously to be active are three families of LTR retrotransposons (*Tos10*, *Tos17*, *Tos19*) that transpose in both *japonica* (Nipponbare) and *indica* (C5924) cell culture¹³. To assess whether *mPing* elements are activated in the same cell lines, we used a modification of the AFLP technique called transposon display^{14,15} to detect *mPing* insertions that might have occurred in culture. Because all of the *mPing* elements were virtually identical at their ends (Fig. 1 and Supplementary Information), we designed element-specific primers located in subterminal sequence (Methods) to amplify all family members and the flanking sequence.

The number of polymerase chain reaction (PCR) products (referred to as amplicons) amplified from DNAs isolated from Nipponbare (*japonica*) and C5924 (*indica*) plants before culture was consistent with the copy-number estimates for these cultivars (Fig. 2a); however, a differential response to cell culture was observed. Whereas the Nipponbare amplicon pattern remained the same, C5924 cells have undergone a marked increase in amplicons. Nonspecific genomic rearrangements during cell culture were ruled out as a cause of the pattern differences by repeating transposon display with primers derived from two other rice transposable elements (a *gypsy*-type LTR retrotransposon (SZ-2; N.J. and S.R.W., unpublished data) and a MITE (*ID-1*; ref. 7). In

contrast to the *mPing* transposon display, these amplicons were essentially identical before and after cell culture.

In several studies, MITEs have been found predominantly in non-coding genic regions^{16–18}. Without an example of an actively transposing MITE, it has not been possible to determine whether this distribution reflects preferential targeting or selection against insertion into other regions of the genome, or whether it reflects the elimination of alleles containing such insertions. To address this issue, we recovered 42 amplicons from cell line C5924 from the transposon display gel and sequenced them. Insertion sites of newly transposed elements were deduced by using sequences flanking the TIR (37–268 bp) to query the 93-11 and Nipponbare sequences. Thirty-four of 42 flanking sequences matched entries from 93-11 contigs, whereas one of the sequences was found only in *japonica* (cv. Nipponbare). Of the 35 matches, 32 were single-copy sequences, and one was a two-copy sequence (Supplementary Table 1). The remaining two insertion sites were in or next to other MITEs that were themselves in single-copy sequences. Thus, 34 of 35 new insertions were in single-copy regions of the genome.

Like other MITE families, *mPing* elements have no coding capacity and, as such, are incapable of catalysing their own transposition. Thus, movement of *mPing* must be catalysed by a transposase encoded *in trans*. To identify putative autonomous elements, we used the *mPing* sequence to query all available rice genomic sequence for related, but longer elements. A single element called *Ping* was found in the Nipponbare but not in the 93-11 draft sequence. *Ping* is 5,341 bp and shares 253 bp and 177 bp, respectively, of its terminal sequences with *mPing*, suggesting that *mPing* arose very recently from *Ping* by internal deletion. Further BLAST searches using *Ping* as the query led to the identification of *Pong* (5,166 bp), with identical 15-bp TIRs and similar subterminal regions to those of *mPing* and *Ping* (Fig. 1). At least five copies of *Pong* were found in Nipponbare and at least three copies in 93-11. Six of these eight copies (five in Nipponbare and one in 93-11) are complete and nearly identical (>99% identity), whereas two are truncated.

In addition to their termini, *Ping* and *Pong* also share similarity in two blocks of internal sequence corresponding to the two main open reading frames (ORFs) of each element (Fig. 1). When used as queries in TBLASTN searches against GenBank databases, both ORFs yielded numerous hits ($E < 10^{-10}$) from several plants, as well as animals and fungi (Supplementary Table 2). ORF2 homologues are

abundant in plants and found most frequently in organisms with large amounts of genomic sequence in databases: 82 hits ($E < 10^{-46}$) from rice, 56 hits ($E < 10^{-23}$) from *Arabidopsis* and over 100 hits ($E < 10^{-36}$) from *Brassica oleracea*. Most ORF1 homologues are located within 2 kb of ORF2 homologues, and several ORF pairs are flanked by TIRs and TSDs that are similar to those of *Ping* and *Pong* (data not shown). It is therefore likely that each 'pair' of ORF1 and ORF2 belongs to the same element.

The function of ORF1 is unclear. It has only very weak sequence similarity to the DNA-binding domains of Myb (Pfam 7.3, $E = 0.002$). By contrast, ORF2 shares significant similarity with the transposase of the maize *PIF* element¹¹, which is also associated with a *Tourist*-like MITE family (called *mPIF*)¹¹. *PIF* belongs to a superfamily of transposons with members identified in all three eukaryotic kingdoms. ORF2 encodes a putative Asp-Asp-Glu (DDE) motif (Fig. 1) containing three acidic amino acids found at the catalytic core of the transposases of bacterial insertion sequences and some eukaryotic transposons. The amino-acid residues surrounding and including the DDE motif (referred to as N2, N3 and C1)¹⁹ are highly conserved among *Ping/Pong*, *PIF* and several IS5-like elements (Fig. 1).

Although *mPing* elements are clearly derived from *Ping*, the evidence suggests that *Ping* is not the autonomous element that mobilizes *mPing* in C5924 cell culture. *Ping* was detected only as a single copy in Nipponbare and it was absent in the draft sequence of 93-11 (~84% of the genome) and in 20 of 24 rice cultivars (8 cultivars for each of the following groups: temperate *japonica*, tropical *japonica* and *indica*) tested by PCR, including C5924 itself. Only four temperate *japonicas* were found to contain *Ping*: Nipponbare, Gihobyoe, JX 17 and Koshikari (data not shown). The apparent absence of *Ping* from all *indica* cultivars tested provides strong evidence that it could not be responsible for the movement of *mPing* elements in the *indica* cell line. *Pong*, by contrast, is present in several near-identical copies in both *indica* and *japonica*. In addition, ORF1 of *Ping* seems to be truncated at the amino terminus as compared with its homologues, and lacks at least 60 conserved amino acids (Fig. 1).

If *Pong* is the autonomous element responsible for the transposition of *mPing* elements, it should also be capable of transposition. By exploiting sequence differences between *Pong* and *mPing*, we designed PCR primers to amplify *Pong* but not *mPing* elements in a transposon display assay. The results with primers for *Pong*

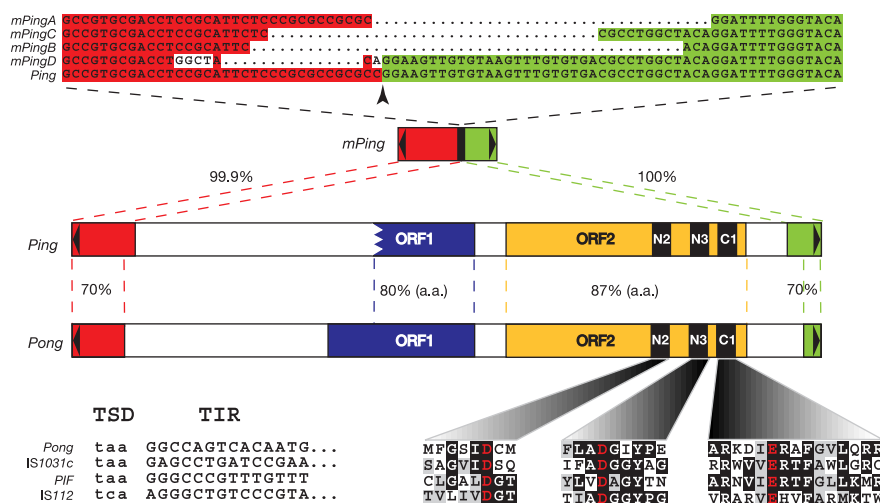


Figure 1 Comparison of *mPing*, *Ping* and *Pong* elements. Black triangles represent TIRs and black boxes represent putative N2, N2 and C1 catalytic domains (nucleotide sequences of the TIRs and TSDs, and amino-acid sequences of the catalytic domains of rice *Pong*, maize *PIF*¹¹ and bacterial IS1031c and IS112 elements are shown). The thick

vertical black line in *mPing* represents internal sequences that differ among the four subtypes derived from *Ping*. An alignment of this region is shown at the top. Arrowhead indicates the breakpoint in *Ping* where 4,923 bp of its internal sequence is not shown in alignment.

mirrored those for *mPing*; that is, the number of amplicons increased markedly in the *indica* cell line but remained virtually the same in Nipponbare (Fig. 2). The nature of the insertion sites and the inserted elements were determined as they were for *mPing*. Nine of ten insertion sites were located in single-copy sequences (Supplementary Table 1). Eight newly inserted elements were successfully amplified by PCR and all were indistinguishable in size from *Pong*.

The difference in the estimated copy number of *mPing* elements in a *japonica* (Nipponbare) and an *indica* (93-11) genome (70 versus 14) suggested the recent amplification of this MITE family, perhaps since domestication. To assess the timing of amplification, we carried out transposon display with a series of *O. sativa* DNAs to determine the approximate copy number of *mPing* and *Pong* elements. The temperate *japonicas* contain the largest number of different *mPing*-anchored amplicons, whereas the tropical *japonicas* contain the least (Fig. 3). This marked difference in *mPing* copy number between the two subgroups of *japonica* is significant because the temperate and tropical cultivars are thought to have diverged since domestication (5,000–7,000 years ago) and are more closely related to each other than either is to *indica*^{20–24}. The different amplicon patterns of *Pong* elements observed in these cultivars also suggest that this element has been active since domestication. But the consistency of amplicon number across cultivars suggests that *Pong* elements have not substantially increased their copy number.

We have reported the use of a systematic computational approach to isolate active transposon elements from a host genome. So far, only a single active DNA transposon has been isolated from a vertebrate (*Tol2*)²⁵ and none has been isolated from mammals. Our

protocol could potentially be applied to the available draft sequences of many organisms, including human, mouse and zebrafish.

With the identification of an active MITE family, long-standing mechanistic issues concerning the birth, spread and death of MITEs can now be addressed experimentally. Some issues have been resolved already; for example, we have shown that the prevalence

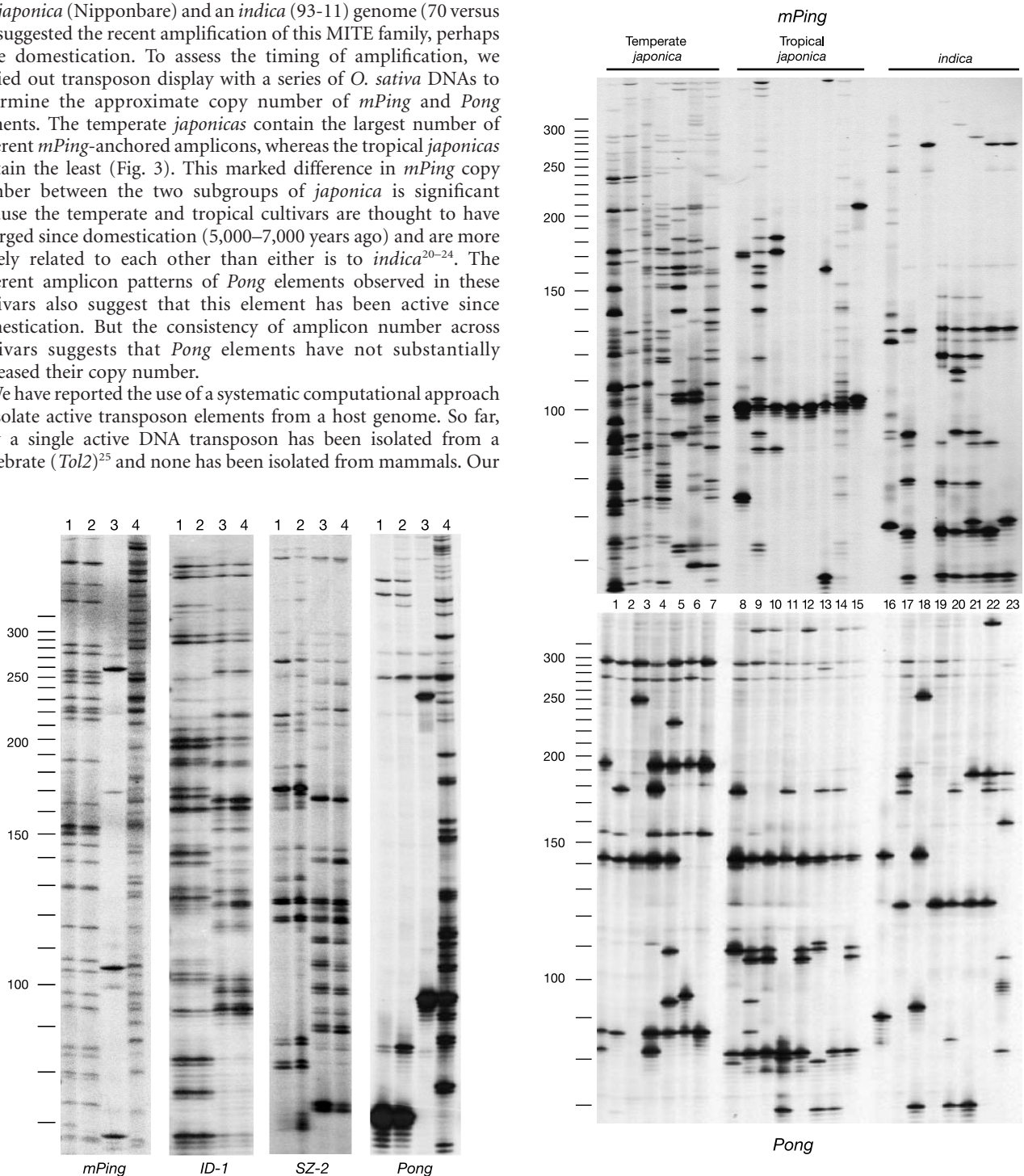


Figure 2 Autoradiograph of transposon display gels of *mPing* and *Pong* amplicons with rice genomic DNAs isolated before and after cell culture. The same genomic DNAs (digested and ligated with adapter primers) were used for each set of primers: 1, Nipponbare; 2, calli of Nipponbare; 3, C5924; 4, Oc cell lines derived from C5924 (ref. 30). The migration of DNA markers is indicated on the left in base pairs.

Figure 3 Autoradiograph of transposon display gels of *mPing* and *Pong*. Genomic DNAs: 1, Nipponbare; 2, Gihobyee; 3, JX 17; 4, Koshikari; 5, Calrose; 6, Early Wataribune; 7, Shinriki; 8, Azucena; 9, Lemont; 10, Jefferson; 11, Moroberekan; 12, Rexoro; 13, Wab56-104; 14, Carolina Gold; 15, Kaybonnet; 16, C5924; 17, IR64; 18, Kasalath; 19, GuangLuAi4; 20, 93-11; 21, Teqing; 22, IR36; 23, Bs125. The migration of DNA markers is indicated on the left in base pairs.

of MITEs in single-copy regions primarily reflects targeting rather than selection. Our results also lead to some unexpected preliminary conclusions. First, *Ping*, which has clearly given rise to *mPing*, seems to be dispensable for the transposition of *mPing* in the C5924 cell line. The fact that in temperate *japonica* cultivars the presence of *Ping* correlates with *mPing* amplification suggests, however, that *Ping* may serve as a coactivator (with *Pong*, perhaps) to enhance transposition of *mPing*. Second, the requirements for transposition of *mPing* may differ in plants and in cell culture. Our data indicate that in cell culture, *Pong*, which is related but distinct from *mPing*, is the most likely source of transposase for this family. This relationship will have to be confirmed experimentally, perhaps in transgenic rice plants. At present, however, our data suggest that one reason for the success of MITEs is an ability to be cross-mobilized by related transposases.

The observation that *mPing* elements seem to have undergone amplification more extensively in temperate than in tropical *japonica* cultivars can be explained by an alternative hypothesis. As mentioned above, temperate and tropical *japonicas* are thought to have diverged from a common ancestor since domestication^{20,21,23}. These two varietal groups are adapted to radically different temperature and water regimes^{24,26,27}. Tropical *japonica* cultivars (previously known as *javanica*) are broadly adapted to tropical and subtropical environments, whereas temperate *japonicas* represent an evolutionary extreme, having been selected for productivity in cool, temperate zones with very short growing seasons. Thus, in a situation reminiscent of the genome shock theory³, stress activation of *mPing* elements during the domestication of temperate *japonicas*, followed by their preferential insertion into genic regions, might have diversified these cultivars and hastened their domestication by creating new allelic combinations that might be favoured by human selection. Notably, it was speculated that an apparent γ -ray-induced insertion in an intron of the rice homologue of the *CONSTANS* gene (*Hd1*) was responsible for a quantitative change in flowering time²⁸. On close inspection, this insertion is identifiable as an *mPing* element. □

Methods

Plant material

DNA from cv. Nipponbare (*japonica*), C5924 (*indica*) and derived cell cultures was obtained from the Hirochika laboratory (National Institute of Agrobiological Resources, Japan). DNA from other rice cultivars was obtained from the McCouch laboratory (Cornell University) or from G. Kochert (University of Georgia). We extracted plant DNA as described²⁹.

Computer-assisted identification of repeat families

Nipponbare sequence (total 187 Mb) was downloaded from <http://rgp.dna.affrc.go.jp> on 24 December 2001, and 93-11 sequence (361 Mb) was downloaded from <http://210.83.138.53/rice/index.php> on 25 February 2002. We used Nipponbare sequence for a systematic identification of repeat families using an all-versus-all comparison with WU-BLASTN2.0 (<http://blast.wustl.edu>; options M=5, N=11, Q=22, R=11, -kap, E=0.0001, -hspmax 5000, -wordmask=dust, -wordmask=seg, -maskextra=50). Alignments were then clustered into repeat families using RECON (<http://www.genetics.wustl.edu/eddy/recon>) with default options and further examined individually using programs from the Genetics Computer Group (GCG; see below). We searched the output (1,257 repeat families) for similarity with features of known MITEs, including short TIRs, TSDs and size (N. Jiang, Z. Bao, S. R. Eddy and S. R. Wessler, manuscript in preparation). Repeat number 1,031 (called *mPing*) was identified as a *Tourist*-like MITE because of its size (430 bp), its TIR similarity to other known *Tourist* MITEs (refs 7, 9, 16 and Fig. 1) and its TSD of TTA or TAA.

Sequences analysis and copy-number estimates

DNA sequence analysis (pairwise comparisons, multiple sequence alignments, sequence assembling and formatting) was done with programs from the University of Wisconsin GCG program suite (version 10.1) accessed through Research Computing Resources at the University of Georgia. Elements related to *mPing* in Nipponbare (updated on 25 February 2002) and 93-11 were identified by BLAST search (WU-BLASTN 2.0) using a derived consensus sequence of *mPing*. Alignment of the central region of *mPing* identified distinct subtypes (*mPingA*, -B, -C and -D; Fig. 1 and Supplementary Table 3). Nipponbare contains only *mPingA*, whereas 93-11 contains *mPingA*, -B and -C; *mPingD* was present only in C5924.

We estimated the copy number of *mPing* as follows: 270 Mb of Nipponbare sequence (including overlaps; downloaded from the Rice Genome Project on 25 April 2002)

contained 44 *mPing* elements ($E < 10^{-70}$) including 36 independent insertions and eight duplicates in regions of overlap. The copy number was estimated to be 70 ($44/270 \times 430$). In 361 Mb of 93-11 contig sequence there were eight complete *mPings* ($E < 10^{-70}$) and four incomplete *mPings* (where the element was located at the end of a contig; $E < 10^{-16}$). Comparison of the flanking sequences of incomplete copies in 93-11 and Nipponbare showed that all incomplete copies were independent insertions. Thus, the copy number in 93-11 was estimated to be 14 ($12/361 \times 430$). Because the 93-11 sequence is in a draft form and was masked before assembly, the fully masked sequence was examined for *mPing*. No *mPing* sequence was found, indicating that the masking process did not significantly affect the assembly of *mPing*. Some *mPing* fragments were found in unassembled sequence but these were not taken into consideration for copy number estimation.

Transposon display and recovery of new insertion sites

Transposon display was done as described¹⁵ with the following modifications. Element-specific primers were designed on the basis of the subterminal sequences of *mPing*, *ID-1*, *SZ-2* and *Pong*, and used in PCR with different rice genomic DNAs. For the display in Fig. 2, the adapter primers were *MseI* + 0 for *mPing* display, *MseI* + AT for *ID-1* display, *BfaI* + C for *SZ-2* display, and *BfaI* + 0 for *Pong* display. For both displays in Fig. 3, the adapter primer was *MseI* + 0. The other primer was specific either for *mPing* (Fig. 3, top) or *Pong* (Fig. 3, bottom). Primer sequences and PCR conditions are given in the Supplementary Methods.

We identified sequences flanking new insertions by cloning and sequencing PCR fragments from transposon display gels as described⁷. The context of the genomic sequence adjacent to the new insertion was determined using a BLAST search (WU-BLASTN 2.0) of the Nipponbare and 93-11 genomic sequence database. Single-copy sequence was defined as a query that results in no more than one hit per genome (except duplicates) with WU-BLASTN 2.0 default parameters.

Search for *Ping/Pong*-related elements

We identified homologues of ORF1 and ORF2 (in *Ping* and *Pong*) by database searches (TBLASTN) using the BLAST server from the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov>) against the non-redundant, expressed-sequence tag and Genome Survey Sequence databases.

Received 11 June; accepted 2 October 2002; doi:10.1038/nature01214.

- Goff, S. A. *et al.* A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* **296**, 92–100 (2002).
- Yu, J. *et al.* A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* **296**, 79–92 (2002).
- McClintock, B. The significances of responses of the genome to challenge. *Science* **226**, 792–801 (1984).
- Arumuganathan, K. & Earle, E. D. Nuclear DNA content of some important plant species. *Plant Mol. Biol. Rep.* **9**, 208–218 (1991).
- Burr, B. Mapping and sequencing the rice genome. *Plant Cell* **14**, 521–523 (2002).
- Tarchini, R., Biddle, P., Wineland, R., Tingey, S. & Rafalski, A. The complete sequence of 340 kb of DNA around the rice *Adh1-Adh2* region reveals interrupted colinearity with maize chromosome 4. *Plant Cell* **12**, 381–391 (2000).
- Jiang, N. & Wessler, S. R. Insertion preference of maize and rice miniature inverted repeat transposable elements as revealed by the analysis of nested elements. *Plant Cell* **13**, 2553–2564 (2001).
- Feschotte, C., Jiang, N. & Wessler, S. R. Plant transposable elements: where genetics meets genomics. *Nature Rev. Genet.* **3**, 329–341 (2002).
- Turcotte, K., Srinivasan, S. & Bureau, T. Survey of transposable elements from rice genomic sequences. *Plant J.* **25**, 169–179 (2001).
- Feschotte, C. & Wessler, S. R. *Mariner*-like transposases are widespread and diverse in flowering genomes. *Proc. Natl Acad. Sci. USA* **99**, 280–285 (2002).
- Zhang, X. *et al.* *P Instability Factor*: an active maize transposon system associated with the amplification of *Tourist*-like MITEs and a new superfamily of transposases. *Proc. Natl Acad. Sci. USA* **98**, 12572–12577 (2001).
- Bao, Z. & Eddy, S. R. Automated *de novo* identification of repeat sequence families in sequenced genomes. *Genome Res.* **12**, 1269–1276 (2002).
- Hirochika, H., Sugimoto, K., Otsuki, Y. & Kanda, M. Retrotransposons of rice involved in mutations induced by tissue culture. *Proc. Natl Acad. Sci. USA* **93**, 7783–7788 (1996).
- Van den Broeck, D. *et al.* Transposon display identifies individual transposable elements in high copy number lines. *Plant J.* **13**, 121–129 (1998).
- Casa, A. M. *et al.* The MITE family heartbreaker (*Hbr*): molecular markers in maize. *Proc. Natl Acad. Sci. USA* **97**, 10083–10089 (2000).
- Bureau, T. E., Ronald, P. C. & Wessler, S. R. A computer-based systematic survey reveals the predominance of small inverted-repeat elements in wild-type rice genes. *Proc. Natl Acad. Sci. USA* **93**, 8524–8529 (1996).
- Zhang, Q., Arbuckle, J. & Wessler, S. R. Recent, extensive and preferential insertion of members of the miniature inverted-repeat transposable element family *Heartbreaker* (*Hbr*) into genic regions of maize. *Proc. Natl Acad. Sci. USA* **97**, 1160–1165 (2000).
- Mao, L. *et al.* Rice transposable elements: a survey of 73,000 sequence-tagged-connectors. *Genome Res.* **10**, 982–990 (2000).
- Rezsos, R., Hallet, B., Delcourt, J. & Mahillon, J. The IS4 family of insertion sequences: evidence for a conserved transposase motif. *Mol. Microbiol.* **9**, 1283–1295 (1993).
- Ting, Y. The origin and evolution of cultivated rice in China. *Acta Agron. Sinica* **8**, 243–260 (1957).
- Glazmann, J. C. Isozymes and classification of Asian rice varieties. *Theor. Appl. Genet.* **74**, 21–30 (1987).
- Wang, Z. Y., Second, G. & Tanksley, S. D. Polymorphism and phylogenetic relationships among species in the genus *Oryza* as determined by analysis of nuclear RFLPs. *Theor. Appl. Genet.* **83**, 565–581 (1992).
- Matsuo, T., Futsuhara, Y., Kikuchi, F. & Yamaguchi, H. *Science of the Rice Plant* (Ministry of Agriculture, Forestry and Fisheries, Tokyo, 1997).
- Morishima, H. & Oka, H. I. Phylogenetic differentiation of cultivated rice. XXVII. Numerical evaluation of the *indica-japonica* differentiation. *Jpn. J. Breed.* **31**, 402–413 (1981).
- Kawakami, K., Shima, A. & Kawakami, N. Identification of a functional transposase of the *tol2*

- element, an *Ac*-like element from the Japanese medaka fish, and its transposition in the zebrafish germ lineage. *Proc. Natl Acad. Sci. USA* **97**, 11403–11408 (2000).
26. Glaszmann, J. C. & Arrau deau, M. Rice plant type variation: 'japonica'–'javanica' relationships. *Rice Genet. Newsl.* **3**, 41–43 (1986).
 27. Ueno, K. Differentiation of ecotypes in *Oryza sativa* L. 2. Characteristics of ecotypes: Japanese lowland and upland rice. *Bull. Inst. Agri. Res. Tohoku Univ.* **39**, 43–49 (1988).
 28. Yano, M. *et al.* *Hd1*, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the *Arabidopsis* flowering time gene *CONSTANS*. *Plant Cell* **12**, 2473–2484 (2000).
 29. McCouch, S. R. *et al.* Molecular mapping of rice chromosomes. *Theor. Appl. Genet.* **76**, 815–829 (1988).
 30. Baba, A., Hasegawa, S. & Syono, K. Cultivation of rice protoplasts and their transformation mediated by *Agrobacterium* *Sphero*plasts. *Plant Cell Physiol.* **27**, 463–471 (1986).

Supplementary Information accompanies the paper on *Nature's* website
(<http://www.nature.com/nature>).

Acknowledgements We thank J. Edwards for help with database searches; C. Feschotte and E. Pritham for critically reading the manuscript; and D. Holligan and Y. Hu for technical assistance. This study was supported by a grant from National Science Foundation to S.R.E., S.R.M. and S.R.W.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for material should be addressed to S.R.W. (e-mail: sue@dogwood.botany.uga.edu). The *mPing*, *Ping* and *Pong* sequences have been deposited in GenBank under accession codes BK000586, BK000587 and BK000588, respectively.

The plant MITE *mPing* is mobilized in anther culture

Kazuhiro Kikuchi*, Kazuki Terauchi†‡, Masamitsu Wada*† & Hiro-Yuki Hirano§

* National Institute for Basic Biology, Okazaki 444-8585, Japan

† Graduate School of Science, Tokyo Metropolitan University, Tokyo 192-0397, Japan

§ Graduate School of Agricultural and Life Sciences, University of Tokyo, Tokyo 113-8657, Japan

Transposable elements constitute a large portion of eukaryotic genomes and contribute to their evolution and diversification. Miniature inverted-repeat transposable elements (MITEs) constitute one of the main groups of transposable elements and are distributed ubiquitously in the genomes of plants and animals¹ such as maize^{2–5}, rice³, *Arabidopsis*^{6,7}, human⁸, insect^{9,10} and nematode¹¹. Because active MITEs have not been identified, the transposition mechanism of MITEs and their accumulation in eukaryotic genomes remain poorly understood. Here we describe a new class of MITE, called *miniature Ping* (*mPing*), in the genome of *Oryza sativa* (rice). *mPing* elements are activated in cells derived from anther culture, where they are excised efficiently from original sites and reinserted into new loci. An *mPing*-associated *Ping* element, which has a putative *PIF* family⁵ transposase, is implicated in the recent proliferation of this MITE family in a subspecies of rice.

Miniature inverted-repeat transposable elements are small transposable elements (usually <500 base pairs), containing short terminal inverted repeats (TIRs), that occur in high copy numbers (~1,000–15,000) in eukaryotic genomes. In plants, MITEs have a preference for target sites that yield 2-bp or 3-bp (A + T)-rich duplications on transposition. Evolutionary studies suggest that some MITEs in plant and animal genomes have spread recently^{1,4,10}; however, molecular evidence for their transposition has not been documented. Doubled haploid breeding, which exploits mutations

that occur with high frequency during anther culture, has been used as an efficient method to improve agronomically important crops such as rice and barley by producing useful cultivars¹². Here we show that a rice MITE is mobilized in anther culture, suggesting that this transposition may account, in part, for the high frequency of mutations in rice.

A transposable-element-like element with TIRs, which we named *mPing*, was identified by database analyses of rice genomic sequence. *mPing* is a short 430-bp element with 15-bp TIRs and lacks an open reading frame (ORF; Fig. 1a, b). Members of this rice-specific MITE family of transposons seem to have a high-preference insertion site because each one is flanked by 3-bp (A + T)-rich duplicated sequences. Of 42 such elements examined, 22 had terminal TAA duplications and 20 had TTA duplications. The copy number of *mPing* in *O. sativa* L. spp. *japonica* was estimated to be 60–80 by database searches (Monsanto database, 60% coverage) and by Southern blot analysis (Fig. 1c). This copy number is unusually low, as the copy number of other MITEs is typically in the thousands^{2–11}. *mPing* sequences are highly conserved: the nucleotide sequences of 34 members (type A1) of this family are completely identical, and the other seven members have only one base change from type A1 members (see Supplementary Information). Southern blot analysis showed that the banding patterns of *mPing* are highly polymorphic among two domesticated rice (*japonica* and *indica*) and their ancestral species, *Oryza rufipogon* (ref. 13 and Fig. 1c).

In contrast to *japonica* rice, which has a pattern with many bands, *indica* rice and *O. rufipogon* have a pattern consistent with fewer *mPing* elements. In agreement with the Southern blot analysis, a database analysis indicated that the *indica* genome (about 92% coverage)¹⁴ contains only eight members of the 430-bp *mPing* MITE family (<http://btn.genomics.org.cn/rice/>). These results indicate

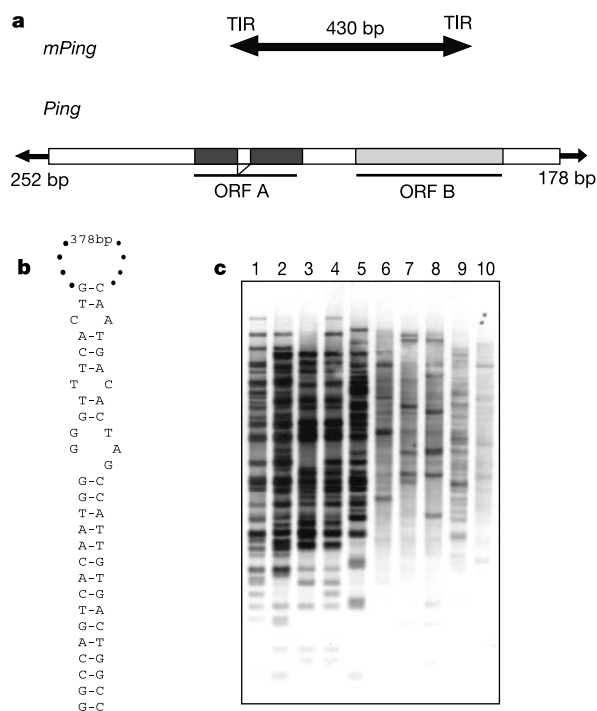


Figure 1 Structures and polymorphisms of *mPing* and *mPing*-related *Ping*. **a**, *mPing* and *Ping* structures. Sequences identical to *mPing* elements are indicated as bold arrows on the *Ping* element. **b**, Sequence of the TIR of *mPing* (complete, 15/15; incomplete 22/27). **c**, Southern blot showing band polymorphism of *mPing* in various rice cultivars and their ancestors. Lanes 1–5, *japonica* cultivars Nipponbare, Sachikaze, Koshihikari, Hitomebore and Taichung 65, respectively; lanes 6–8, *indica* cultivars Kasalath, IR36 and T0710; lanes 9 and 10, *O. rufipogon* W1987 and W0125.

‡ Present address: Graduate School of Science, Osaka University, Toyonaka, Osaka 560-0043, Japan.

that *mPing* was amplified in the *japonica* lineage after its divergence from *O. rufipogon*. In addition, polymorphic bands were detected even in related *japonica* cultivars (Fig. 1c), suggesting that *mPing* is currently mobile. It should be noted that two pairs of *japonica* cultivars, Sachikaze–Nipponbare and Koshihikari–Hitomebore, were in parent–child relationships in breeding programs.

The rice retrotransposon *Tos17* is mobilized under conditions of tissue culture¹⁵. To test the mobilization of *mPing*, we first examined the presence or absence of *mPing* members at five loci (J01, J03, J05, J11, J12) by polymerase chain reaction (PCR), using 64 independent calli that were induced from the anthers and scutellum of *japonica* rice (Fig. 2a and Table 1). The presence of both upper and lower bands in one lane of Fig. 2a indicated that these anther-derived calli were composed of heterogeneous cells, in other words, cells carrying *mPing* and cells not carrying *mPing*. The different intensities of the two bands implied differences in the timing of excision during cell proliferation (Fig. 2a). These results indicate that *mPing* elements were excised from the original site at a high frequency in anther-derived calli, but that their excision was very low in scutellum-derived calli (Table 1). Thus, activation of *mPing* occurs preferentially in cells derived from anthers, suggesting that genes that function in germ cells are required for this activation. The excision also indicates that *mPing*, a transposon in the MITE family, belongs to class 2 transposable elements and transposes by a cut-and-paste mechanism.

In addition to *mPing* elements, database searches of the *japonica* genome also identified a larger element, termed *Ping*. In the *Ping* element, the *mPing* sequence is separated into a 252-bp left part and

Table 1 Excision frequency of *mPing* and *Ping*

	<i>mPing</i> loci					<i>Ping</i>
	J01	J03	J05	J11	J12	
Calli derived from scutellum						
Number of excisions	0	1	1	1	0	0
Frequency (%)	0	1.6	1.6	1.6	0	0
Calli derived from anther						
Number of excisions	7	20	17	14	11	4
Frequency (%)	10.9	31.3	26.6	21.9	17.2	6.3

a 178-bp right part by a 4,911-bp sequence that contains two putative ORFs (Fig. 1a). The terminal sequences of *Ping* are almost identical (429/430) to those of *mPing*. *Ping*, like *mPing*, is flanked by TTA duplications (Fig. 1b), suggesting that it is also a mobile element. Indeed, the excision of *Ping* was also detected in anther-derived calli (Table 1).

Sequencing of the excision indicated that almost all excisions of *mPing* in anther culture effected changes in the surrounding region (Fig. 2b). The 5' and/or 3' flanking sequences were mostly deleted (Fig. 2b, J12), but sometimes additional DNA sequences (known as filler DNAs¹⁶) were inserted into the site where *mPing* had been present originally. The original location of these filler DNAs were 57-bp upstream of the *mPing* insertion site in J01-d39 (Fig. 2b) and J01-d24 (data not shown), and 162-bp upstream in J01-d54 (data

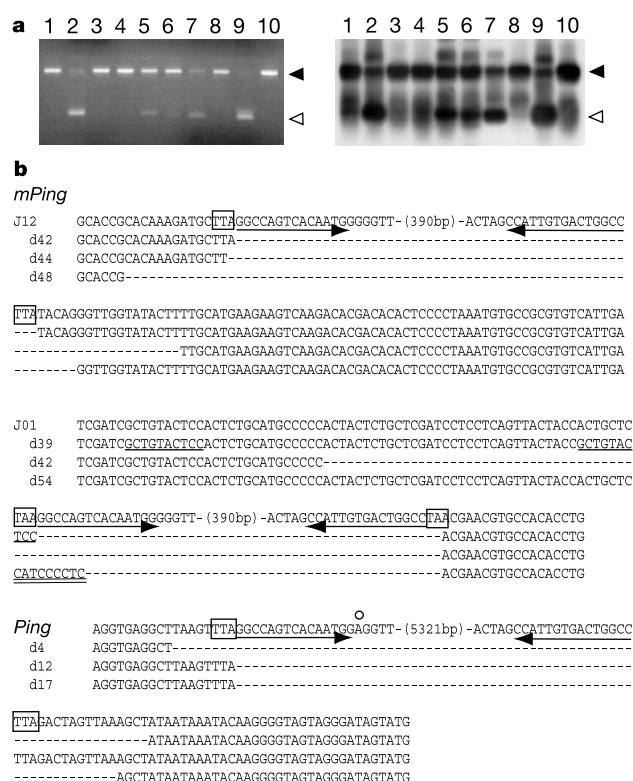


Figure 2 Excision of *mPing* and *Ping*. **a**, An example of *mPing* excision (J01) in anther-derived calli. Left, ethidium bromide staining; right, Southern blot. Filled triangle indicates bands including *mPing*; open triangle indicates bands lacking it. **b**, Footprints after excision of *mPing* and *Ping*. Sequences after their excision are shown with callus number (d4, d12, d17, d39, d42, d44, d48 and d54). TIRs are indicated with arrows and target-site duplications (TSDs) are boxed. Filler DNA is indicated by an underline (together with its original sequence) and with a double underline. Open circle in *Ping* denotes a base that differs from *mPing*.

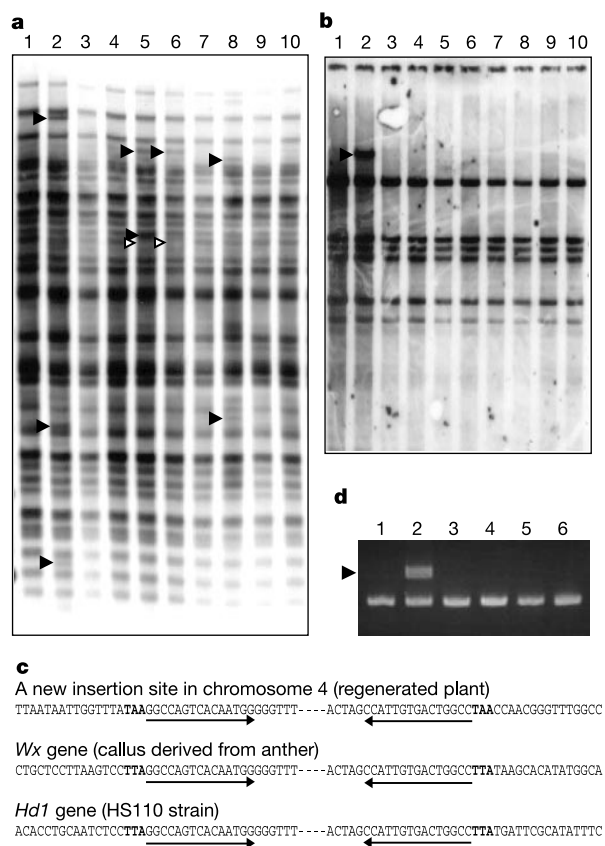


Figure 3 Transposition of *mPing* and *Ping*. **a**, Southern blot analysis of Nipponbare (lane 1) and nine plants regenerated from anther-derived calli (lanes 2–10) using *mPing* as a probe. Filled triangles indicate newly appearing bands; open triangles indicate deleted bands (excision). **b**, Same as in **a**, except the probe was a part of ORF B in *Ping*. **c**, Sequences at the *mPing* insertion sites. The top sequence is from the regenerated plant analysed in lane 2 of panel **a**. The TIRs are indicated with arrows and the TSDs are indicated with bold letters. **d**, PCR amplification of DNA at a new insertion site. Filled triangle indicates a band including *mPing*. The lanes are the same as in **a**.

not shown). The excision of *Ping* also produced a footprint at the deletion site. In one case (d12), *Ping* was excised without deletion of the surrounding DNA, leaving only a single TAA sequence.

To examine whether *mPing* and *Ping* sequences had been integrated into new positions, rice plants were regenerated from anther-derived calli and total genomic DNA was analysed by Southern blot. Fragments hybridizing to *mPing* showed polymorphic banding patterns in different isolates (Fig. 3a). In addition to disappearance of a few bands (open triangles), several new bands (filled triangles) were detected in some regenerated plants, suggesting that *mPing* had reinserted into new positions. A new insertion band was also detected in *Ping* (Fig. 3b). To confirm the reinsertion of *mPing*, we amplified DNA fragments by inverse PCR using genomic DNA from a regenerated plant (Fig. 3a, lane 2) and cloned them. By sequencing 120 clones chosen at random, we identified an *mPing* sequence inserted into a new locus (chromosome 4). The sequence of this new insertion was identical to that of *mPing* (type A1) and was flanked by duplications of a typical 3-bp sequence (TAA; Fig. 3c). PCR amplification of this locus revealed that only this regenerated plant had the *mPing* insertion (Fig. 3d). These results strongly support the idea that *mPing* had been integrated into a new position.

We next used PCR to amplify genomic DNA of more than 600 calli derived from anther culture and gene-specific primers for roughly 20 genes, including the *Waxy* (*Wx*) and *OsNAC* loci^{17,18}. We found that *mPing* (type A1) had been integrated into the second exon of the rice *Wx* gene (Fig. 3c). In addition, a type A1 *mPing* sequence was identified by BLAST search in the mutant allele of the *Hd1* locus (HS110)¹⁹, which had been derived from the strain Ginbozu by γ -ray irradiation (Fig. 3c). The same rice MITE has been shown to be mobile and to give rise to mutable seed phenotypes²⁰. Together, these results strongly support the hypothesis that *mPing* transposition is activated by stresses such as anther culture and γ -rays.

With regard to the transposition of *mPing* sequences, it is tempting to speculate that the *Ping* element is responsible for this activity. Indeed, our studies show that *Ping* is active in cells derived from anther cultures of *japonica* rice. We also found that ORF B in *Ping* has partial homology to putative transposases in the *PIF* family (ref. 5 and Supplementary Information). In addition, none of the strains of *indica* and *O. rufipogon* that were examined had a *Ping* sequence (data not shown). This might account for the lower number of *mPing* elements in these species than in *japonica* rice. These results raise the possibility that *Ping* is a putative autonomous element from which *mPing* sequences are derived. In this regard, *mPing* is similar to maize *miniPIF*⁵, *Arabidopsis Emigrant*²¹ and human MER⁸, which are thought to be derived from larger autonomous-like elements.

The fact that the copy number of *mPing* (60–80) even in *japonica* rice is much lower than those of other MITEs (1,000s)^{2–11} suggests that the amplification of *mPing* may be currently at an early stage of MITE accumulation, because *japonica* rice has evolved from *O. rufipogon* in a relatively short time period (at least ten thousand years). Activation of *mPing* and *Ping* in cells from anther culture indicates that the transposition of these elements may be one of the sources of mutations caused by anther culture in rice, similar to the way in which *Tos17* is one of the causes of somatic variation in rice¹⁵. Because *mPing* seems to be inserted in the genic region like other plant MITEs, it should be a useful molecular tool for gene isolation and gene knockouts in these important crop plant species. □

Methods

Plant materials and sequence information

We obtained rice strains from the Genetic Strains Research Center of the National Institute of Genetics and the MAFF Genebank of the National Institute of Agrobiological Sciences in Japan. The sequence databases used in this study were the DNA Data Bank of Japan (DDBJ) public database as of March 2002 and a private database²² of Monsanto (<http://www.rice-research.org/>). We identified 40 complete elements of *mPing*, which are located

in independent loci in the Pharmacia *O. sativa* L. spp. *japonica* database. Because this database covers about 60% of the rice genome, we estimated the copy number of *mPing* to be 60–80.

Culture methods and detection of transposition

O. sativa L. spp. *japonica* cv. Nipponbare was used for the scutellum and anther culture. Callus induction from the scutellum (10 d) and subsequent cell propagation (14 d) in fresh medium were carried out essentially as described²³. Calli were induced from anthers for 4 weeks and cells were propagated for 2 weeks in a fresh medium as described²⁴ using anthers collected from panicles before heading. We extracted genomic DNA from calli and used it for PCR amplification. Primers used to detect the excision of *mPing* by PCR amplification were *mPing* J01, 5'-CACAAGTGGTTTATGGTTAGC-3' and 5'-AATTGTTTCGAGAGT TTTAGT-3'; J03, 5'-TCCATCCTCCTACTCCTCCACA-3' and 5'-GATCCATAACTGCC TACAAAGA-3'; J05, 5'-ATGTAGTTTGTCCGTAAGTTGA-3' and 5'-ATGTGTTGTG ATTGATGGGATAA-3'; J11, 5'-TTCCGGTGGTTGTGTTTGTGCTTGC-3' and 5'-AGGGTGTGGTTGTGGTTGTGGTAG-3'; J12, 5'-GTGCGTGGTTGGTCTCGGCTT TAT-3' and 5'-CCTCCTCTCCTCCTGGTGTGG-3'; *Ping* 5'-GGGTGAGTGAAGTG AGTGAGTGAGCAGCAT-3' and 5'-AGTTAGGGGAGGAGAGTTGGGCATAGGAAT-3'.

For Southern blot analysis, genomic DNA was digested with *Hind*III, *Xba*I and *Xho*I (sites not found in *mPing*), separated on a 1% agarose gel, and subjected to hybridization at high stringency using the probes indicated in the figure legends. To detect new insertion sites, genomic DNA from a regenerated plant (Fig. 3a, lane 2) was digested with *Alu*I (sites that are not found in *mPing*), self-ligated and amplified with a primer (5'-CCATTGTT GACTGGCC-3') in the TIR (inverse PCR). To verify insertion, we carried out PCR amplification of a new insertion site using same genomic DNAs as those used in Southern analysis (Fig. 3a) and the primers 5'-ATTGTTTGTGTTGTTGTTGTTGGAG GGCTAA-3' and 5'-CATATGCTGGTTGGAGACTTGGACGACTAA-3'.

Received 28 June; accepted 3 October 2002; doi:10.1038/nature01218.

- Feschotte, C., Jiang, N. & Wessler, S. R. Plant transposable elements: where genetics meets genomics. *Nature Rev. Genet.* **3**, 329–341 (2002).
- Bureau, T. E. & Wessler, S. R. *Tourist*: A large family of small inverted repeat elements frequently associated with maize genes. *Plant Cell* **4**, 1283–1294 (1992).
- Bureau, T. E. & Wessler, S. R. Mobile inverted-repeat elements of the *Tourist* family are associated with the genes of many cereal grasses. *Proc. Natl Acad. Sci. USA* **91**, 1411–1415 (1994).
- Zhang, Q., Arbuckle, J. & Wessler, S. R. Recent, extensive, and preferential insertion of members of the miniature inverted-repeat transposable element family *Heartbreaker* into genic regions of maize. *Proc. Natl Acad. Sci. USA* **97**, 1160–1165 (2000).
- Zhang, X. *et al.* *P* instability factor: an active maize transposon system associated with the amplification of *Tourist*-like MITEs and a new superfamily of transposases. *Proc. Natl Acad. Sci. USA* **98**, 12572–12577 (2001).
- Surzycki, S. A. & Belknap, W. R. Characterization of repetitive DNA elements in *Arabidopsis*. *J. Mol. Evol.* **48**, 684–691 (1999).
- Casacuberta, E., Casacuberta, J. M., Puigdomenech, P. & Monfort, A. Presence of miniature inverted-repeat transposable elements (MITEs) in the genome of *Arabidopsis thaliana*: Characterization of the *Emigrant* family of elements. *Plant J.* **16**, 79–85 (1998).
- Smit, A. F. A. & Riggs, A. D. *Tiggers* and other DNA transposon fossils in the human genome. *Proc. Natl Acad. Sci. USA* **93**, 1443–1448 (1996).
- Tu, Z. Molecular and evolutionary analysis of two divergent subfamilies of a novel miniature inverted repeat transposable element in the yellow fever mosquito *Aedes aegypti*. *Mol. Biol. Evol.* **17**, 1313–1325 (2000).
- Feschotte, C. & Mouches, C. Recent amplification of miniature inverted-repeat transposable elements in the vector mosquito *Culex pipiens*: Characterization of the *Mimo* family. *Gene* **250**, 109–116 (2000).
- Surzycki, S. A. & Belknap, W. R. Repetitive-DNA elements are similarly distributed on *Caenorhabditis elegans* autosomes. *Proc. Natl Acad. Sci. USA* **97**, 245–249 (2000).
- Raina, S. K. Doubled haploid breeding in cereals. *Plant Breed. Rev.* **15**, 141–186 (1997).
- Morishima, H., Sano, Y. & Oka, H.-I. Evolutionary studies in cultivated rice and its wild relatives. *Oxf. Surv. Evol. Biol.* **8**, 135–184 (1992).
- Yu, J. *et al.* A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* **296**, 79–92 (2002).
- Hirochika, H., Sugimoto, K., Otsuki, Y., Tsugawa, H. & Kanda, M. Retrotransposons of rice involved in mutations induced by tissue culture. *Proc. Natl Acad. Sci. USA* **93**, 7783–7788 (1996).
- Wessler, S., Tarpley, A., Purugganan, M., Spell, M. & Okagaki, R. Filler DNA is associated with spontaneous deletions in maize. *Proc. Natl Acad. Sci. USA* **87**, 8731–8735 (1990).
- Hirano, H.-Y., Eiguchi, M. & Sano, Y. A single base change altered the regulation of the *Waxy* gene at the post-transcriptional level during domestication of rice. *Mol. Biol. Evol.* **15**, 978–987 (1998).
- Kikuchi, K. *et al.* Molecular analysis of the NAC gene family in rice. *Mol. Gen. Genet.* **262**, 1047–1051 (2000).
- Yano, M. *et al.* *Hd1*, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the *Arabidopsis* flowering time gene *CONSTANS*. *Plant Cell* **12**, 2473–2483 (2000).
- Nakazaki, T. *et al.* Mobilization of a transposon in the rice genome. *Nature* this issue.
- Feschotte, C. & Mouches, C. Evidence that a family of miniature inverted-repeat transposable elements (MITEs) from the *Arabidopsis thaliana* genome has arisen from a *pogo*-like DNA transposon. *Mol. Biol. Evol.* **17**, 730–737 (2000).
- Barry, G. F. The use of the Monsanto draft rice genome sequence in research. *Plant Physiol.* **125**, 1164–1165 (2001).
- Hiei, Y., Ohta, S., Komari, T. & Kumashiro, T. Efficient transformation of rice (*Oryza sativa* L.) mediated by *Agrobacterium* and sequence analysis of the boundaries of the T-DNA. *Plant J.* **6**, 271–282 (1994).
- Negishi, T., Fujimura, T. & Sakurai, M. The efficient method of production of rice haploid-plants. Japanese Patent 06-036698 B (1994).

Supplementary Information accompanies the paper on *Nature's* website (<http://www.nature.com/nature>).

Acknowledgements We thank T. Tanisaka for providing information on *Sairyu* before publication; T. Fujimura for technical advice on anther culture; S. Iida for critically reading the manuscript; and the Monsanto Rice-Research.Org Program for making available a draft of *japonica* genomic sequences.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for material should be addressed to H.-Y.H. (e-mail: ahirano@mail.ecc.u-tokyo.ac.jp). Sequences of *mPing* and *Ping* have been deposited in the DDBJ under accession numbers AB087615 and AB087616, respectively.

Mobilization of a transposon in the rice genome

Tetsuya Nakazaki*, Yutaka Okumoto*, Akira Horibata†, Satoshi Yamahira*, Masayoshi Teraishi*, Hidetaka Nishida*, Hiromu Inoue* & Takatoshi Tanisaka*

* Division of Agronomy and Horticulture Science, Graduate School of Agriculture, Kyoto University, Kitashirakawa, Sakyo-ku, Kyoto 606-8502, Japan

† Faculty of Biology-Oriented Science and Technology, Kinki University, Uchida, Wakayama 649-6433, Japan

Rice (*Oryza sativa* L.) is an important crop worldwide and, with the availability of the draft sequence^{1,2}, a useful model for analysing the genome structure of grasses^{3,4}. To practice efficient rice breeding through genetic engineering techniques, it is important to identify the economically important genes in this crop. The use of mobile transposons as gene tags in intact plants is a powerful tool for functional analysis because transposon insertions often inactivate genes⁵. Here we identify an active rice transposon named *miniature Ping* (*mPing*) through analysis of the mutability of a slender mutation of the glume⁶—the seed structure that encloses and determines the shape of the grain. The *mPing* transposon is inserted in the *slender glume* (*slg*) mutant allele but not in the wild-type allele. Search of the

O. sativa variety Nipponbare genome identified 34 sequences with high nucleotide similarity to *mPing*, indicating that *mPing* constitutes a family of transposon elements. Excision of *mPing* from *slg* plants results in reversion to a wild-type phenotype. The mobility of the transposon *mPing* in intact rice plants represents a useful alternative tool for the functional analysis of rice genes.

A mutable slender glume mutant (mutant line IM294) was induced by γ -ray irradiation of seeds of the *japonica* rice variety Gimbozu⁶. This mutant character is governed by a recessive mutant allele *slender glume* (*slg*) at the *slg* locus and has not been fixed in spite of repeated self-propagation. In successive generations, both normal glume (wild-type) plants and plants chimaeric for glume shape (see Supplementary Information) occur at a low frequency. *slg* is located in a region covered by four overlapping yeast artificial chromosome (YAC) clones on chromosome 7 (ref. 7).

To pinpoint the genomic location of *slg*, we searched for loci in the overlapping region of two of the YAC clones Y1774 and Y3356. Consequently, we identified a single-copy restriction fragment length polymorphism locus *R430* (ftp://rgp.dna.affrc.go.jp/pub/geneticmap98/parentsouthern/chr07/R430.JPG, last updated on 19 January 2001) from the public data of the Rice Genome Research Program (RGP), Japan (http://rgp.dna.affrc.go.jp/publicdata/geneticmap2000/index.html, last updated on 19 January 2001), and the Integrated Rice Genome Explorer (INE, http://rgp.dna.affrc.go.jp/giot/INE.html, last updated on 20 December 2000). By carrying out a BLAST search with a partial complementary DNA sequence, D23858, found in the *R430* locus, we identified four cDNAs sequences, C73435, C74027, AU101270 and AU164540, in the *R430* locus. In polymerase chain reaction (PCR) analyses with U1 and L1 primers, which were designed by sequencing nucleotides 2–655 of AU101270 (Fig. 1b), IM294 and Gimbozu both yielded a single PCR product, but the former product was larger than the latter (Fig. 1a). This suggested that IM294 and Gimbozu have different alleles at the *R430* locus.

Sequencing analysis of PCR products showed that the Gimbozu allele contains three introns and an open reading frame (ORF) with the same sequence as AU101270 (Fig. 1b). The deduced amino-acid sequence of the Gimbozu allele was not homologous to any plant proteins, but shared homology (41% identity) with the product of

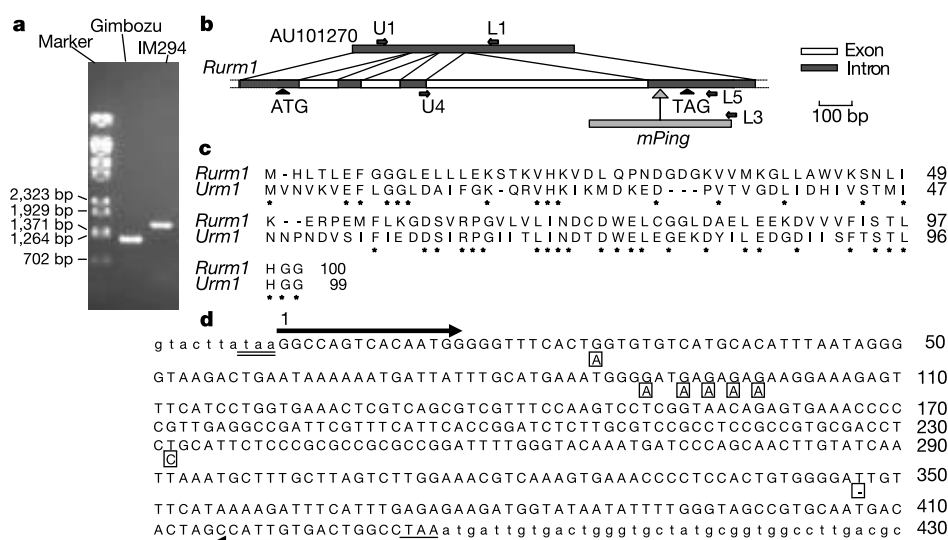


Figure 1 Sequence structure of *Rum1⁺* in Gimbozu and *Rum1^m* in IM294. **a**, PCR products amplified by the U1 and L1 primer pair. **b**, Structural differences between *Rum1^m* and *Rum1⁺*. *Rum1^m* contains the 430-bp insertion sequence '*mPing*'. Arrows indicate the positions and directions of the primers. **c**, Deduced amino-acid sequences of *Rum1* and *Urm1* (GenBank accession number P40554). Asterisks and bars

indicate, respectively, identical residues and gaps between the two sequences. **d**, *mPing* sequence in *Rum1^m*. Lower-case letters, double-underlines and arrows indicate the *mPing* flanking region, TSDs and TIRs, respectively. Bases shown under the *mPing* sequence indicate those found in other *mPing* sequences detected in the Nipponbare genome (Table 1).

Table 1 *mPing* sequence variations in the Nipponbare genome*

Type	Number of sequences	Target-site duplication	Number of base changes	Sequence differences from the <i>Rurm1^m</i> allele in IM294 Nucleotide changes (location in <i>mPing</i> of <i>Rurm1^m</i>)
A	2	TAA	0	
B	2	TTA	0	
C	16	TAA	1	T to C (232)
D	11	TTA	1	T to C (232)
E	1	TAA	2	G to A (26), T to C (232)
F	1	TTA	2	T to C (232), T to null (347)
G	1	TTA	6	T to C (232), G to A (90, 93, 95, 97, 99)

* Accession numbers of sequences including *mPing* are given in the Supplementary Information.

the yeast gene *ubiquitin-related modifier-1* (*urm1*; ref. 8 and Fig. 1c). We therefore named the *R430* locus *rice ubiquitin-related modifier-1* (*Rurm1*), and designated the Gimbozu allele as *Rurm1⁺* and the IM294 allele as *Rurm1^m*.

We found that the two alleles had the same nucleotide sequence except that *Rurm1^m* had a 433-bp insertion sequence in exon 4 (Fig. 1b). The insertion sequence included 15-bp terminal inverted repeats (TIRs) and a 3-bp target-site duplication (TSD) of TAA (Fig. 1d), which suggested that it belonged to the *Tourist*-like superfamily of miniature inverted-repeat transposable elements (MITEs)^{9,10}. But its TIR was not homologous to that of any of the known *Tourist*-like elements (ref. 11 and <http://www.tebureau.mcgill.ca/rice/main.html>).

In a BLAST search carried out in April 2002, we identified 34 sequences with high nucleotide similarity to the insertion sequence

in the genome of Nipponbare, a *japonica* rice variety that is closely related to Gimbozu (Table 1). The TSDs of these sequences formed two groups: TAA, and its complementary sequence TTA. In addition to these similarities in the TSD, four of the 34 sequences were completely identical to the insertion sequence of *Rurm1^m*; 27 differed from it by a single base (T to C) at the same position; and 3 differed from it at more than two bases (Table 1). Thus, these sequences including the insertion sequence constitute a previously unknown family of MITEs, which we called '*mPing*'.

To investigate the relationship between *Rurm1* and the *slg* locus, we carried out PCR analysis using the U1 and L1 primers on 15 normal glume progeny lines (*M₅* and *M₆*, *slg⁺/slg⁺*) derived from three revertants (*M₁* plants, *slg⁺/slg*) in IM294. All of the selfed-progeny lines (*M₅* or *M₆*) of the *slg⁺/slg⁺* homozygous plants (*M₄* or *M₅*) yielded only a *Rurm1⁺*-type PCR product, whereas two lines of the heterozygous plants (*slg⁺/slg*) yielded *Rurm1⁺*- and *Rurm1^m*-type products (Fig. 2). The sequences of the *Rurm1⁺* type (Fig. 2) were all identical to that of Gimbozu. This indicates that *slg* is the same locus as *Rurm1*, and that the reversion of slender glume to its wild type results from the precise excision of *mPing* from the *Rurm1^m* allele.

We analysed this further using an *F₂* population consisting of 3,697 plants from a cross between IM294 and the normal glume line ML17 (ref. 6), which lacks the *mPing* insertion at the *Rurm1* locus. Of the 479 slender glume segregants, 472 yielded only a *Rurm1^m*-type PCR product, whereas seven yielded *Rurm1^m*- and *Rurm1⁺*-type products. Sequencing analysis showed that these exceptional *Rurm1⁺*-type products did not include an *mPing* insertion, but had various kinds of deletions and insertions at the original site of *mPing* (Fig. 3). This implies that *mPing* was excised imprecisely. Thus, these exceptional plants were not recombinants between the two putatively different loci *slg* and *Rurm1*.

Next, we attempted to confirm the occurrence of somatic excision of *mPing* by analysing by PCR a mixed-panicle chimaeric plant with five slender panicles and one normal panicle. Leaves taken from a stem with the normal glume panicle yielded *Rurm1⁺*-type and *Rurm1^m*-type PCR products, whereas those on stems with slender glume panicles yielded only the *Rurm1^m*-type. Thus, the genotype of each stem was completely consistent with that of glume shape. This confirms that the precise somatic excision of *mPing* causes the chimaeras for glume shape.

Most families of MITEs are present in a plant genome at a high copy number (>1,000)¹². From the number of DNA fragments amplified by the transposon display technique^{13–16} on *mPing* (data not shown), we estimated the copy number of *mPing* to be less than

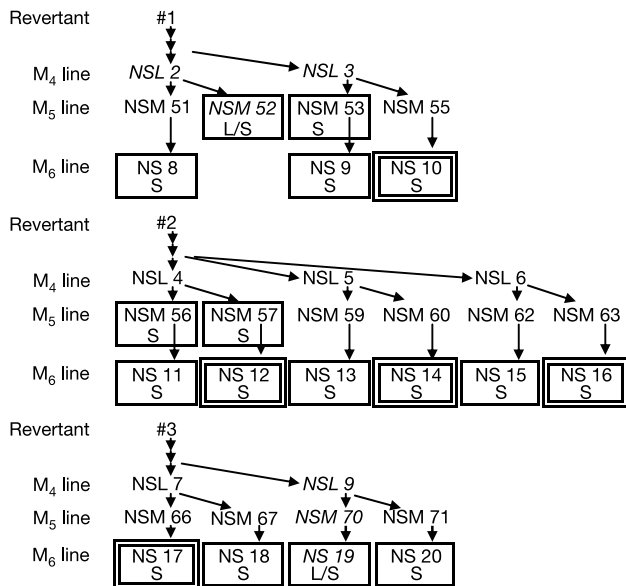


Figure 2 Pedigrees of *M₅* and *M₆* lines used for investigating the relationship between two putatively different loci, *Rurm1* and *slg*. These progenies were derived from three revertants (*M₁*) in IM294. Progenies shown in italics are segregating lines for the *slg* locus; the other progenies are fixed lines. Progenies shown in single boxes were investigated for the genotype of the *Rurm1* locus (S, *Rurm1⁺/Rurm1⁺*; S/L, *Rurm1⁺/Rurm1^m*) by PCR analysis. The PCR products of progenies indicated in double boxes were sequenced.

IM294	TGGTGTCTCTCGTACTTATAA	<i>mPing</i> (430 bp)	TAAATGATTGTGACTGGGAGCTATGCGG
Gimbozu	TGGTGTCTCTCGTACTTATAA		ATGATTGTGACTGGGAGCTATGCGG
3C6	TGGTGTCTCTCGTACTTATAA	A	ATGATTGTGACTGGGAGCTATGCGG
13E5	TGGTGTCTCTCGTACTT...	GACGCGGAGTT	ATGATTGTGACTGGGAGCTATGCGG
20G3	TGGTGTCTCTCGTACTT...	GACGC	ATGATTGTGACTGGGAGCTATGCGG
511	TGGTGTCTCTCGTACTT...	TTTTCTGTGCTT	ATGATTGTGACTGGGAGCTATGCGG
22A4	TGGTGTCTCTCGTACTT...	GAAGG	ATGATTGTGACTGGGAGCTATGCGG
31C6	TGGTGTCTCTCGTACTT...		ATGATTGTGACTGGGAGCTATGCGG
14C1	TGGTGTCTCTCGTACTT...		ATGATTGTGACTGGGAGCTATGCGG

Figure 3 Various kinds of imprecise excision of *mPing* insertion. These types of excision were observed in seven of 479 slender glume segregants (plant numbers 3C6 to 14C1) in

the *F₂* population from the cross between IM294 (*slg/slgl*) and ML17 (*slg⁺/slg⁺*). Deleted bases are indicated by bars; inserted bases are shown under the *mPing* region in the box.

100 (~80) in the IM294 genome. In addition, *mPing* has a lower copy number in *indica* subspecies than in *japonica* subspecies (K. Kikuchi, K. Terauchi, M. Wada, M. and H.-Y. Hirano, personal communication). Therefore, *mPing* is a newly originated MITE family in the rice genome whose amplification began after differentiation into the *japonica* and *indica* subspecies. The low copy number of *mPing* will be a great advantage in efficient gene-tagging procedures for detecting newly inserted sequences in the rice genome. The retrotransposon *Tos17* becomes active under tissue-culture conditions and has been used to develop a large-scale series of rice mutants in the RGP^{17–19}. Although this mutant series is a useful tool for analysing gene functions, its gene-tagging efficiency is very low (5–10%) owing to mutations induced by other factors under tissue-culture conditions²⁰. In addition, tissue-culture techniques for *indica* subspecies have not been fully developed. Thus, the mobility of the *mPing* transposon in intact rice plants will provide a useful alternative tool for analyses based on reverse genetics in both the *indica* and *japonica* subspecies. □

Methods

DNA extraction, PCR and sequencing

We extracted total genomic DNAs from fresh leaves by the phenol–chloroform method²¹, and carried out PCR using rTaq DNA polymerase (Toyobo) and a PCR Thermal Cycler TP 2000 (Takara Biomedicals). The PCR conditions were as follows: pre-denaturation for 5 min at 94 °C followed by 30 cycles of polymerization reaction, each consisting of a denaturation step for 1 min at 94 °C, an annealing step for 2 min at 58 °C and an extension step for 3 min at 72 °C, with a final extension step for 7 min at 72 °C. PCR products were resolved on 0.8% agarose gels with 0.5 × TBE buffer and sequenced by the DYEnamic ET Terminator Cycle Sequencing kit (Amersham Pharmacia) on an ABI 310 Genetic Analyzer (Applied Biosystems). To identify the genotype for the *Rurm1* locus of each line or plant, we carried out PCR using the following primer pair: U1 (5'-CTTCGCCTCCTCCAA TCCT-3') and L1 (5'-AAAACAACAACATCTTCTCCTC-3'). For sequencing *Rurm1* alleles of Gimbozu and IM294 (Fig. 1b), we obtained PCR products using the additional primer pairs U1 and L5 (5'-CATAACTGACAAGTGCCTTTTC-3'), U4 (5'-CAGTGT ATGATTTCCCTTTCC-3') and L1, and U4 and L3 (5'-CTCCAGTCACAATCATT TAGG-3').

Transposon display

For the digestion of genomic DNAs and ligation of a *MseI* adapter primer, we used *MseI* (BioLabs) and Ligation high (Toyobo), respectively. Pre-amplification was done by using an *mPing*-specific primer (5'-GTGGGATGTTTTCATAA-3') and an adapter-specific primer (*MseI*-APO; 5'-GACGATGAGTCCTGAGTAA-3'). For 'hot' amplification, another *mPing*-specific primer (5'-TCATTGAGAGAAGATGGTA-3') labelled by D2-PA (Beckman Coulter) and a primer with an additional base at the 3' end of the primer *MseI* (APO) were used. Labelled *mPing* flanking fragments were analysed by the Capillary DNA Sequencer CEQ2000XL (Beckman Coulter). PCR was done by using Polymerase KOD-Plus (Toyobo) and the Program Temp Control System PC-707 (Astec).

Cultivation of plants

We grew the plants in a paddy field in Kyoto, Japan. We observed glume shape at maturity.

Received 28 June; accepted 3 October 2002; doi:10.1038/nature01219.

1. Yu, J. *et al.* A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* **296**, 79–92 (2002).
2. Stephen, A. *et al.* A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* **296**, 92–100 (2002).
3. Delseny, M. *et al.* Rice genomics: present and future. *Plant Physiol. Biochem.* **39**, 323–334 (2001).
4. Cartinhour, S. W. Public informatics resources for rice and other grasses. *Plant Mol. Biol.* **35**, 241–251 (1997).
5. Kunze, R., Seadler, H. & Lönning, W. Plant transposable elements. *Adv. Bot. Res.* **27**, 321–470 (1997).
6. Teraishi, M. *et al.* Identification of a mutable slender glume gene in rice (*Oryza sativa* L.). *Mol. Gen. Genet.* **261**, 487–494 (1999).
7. Teraishi, M. *et al.* Identification of YAC clones containing the mutable slender glume locus in rice (*Oryza sativa* L.). *Genome* **44**, 1–6 (2001).
8. Furukawa, K., Mizushima, N., Noda, T. & Ohsumi, Y. A protein conjugation system in yeast with homology to biosynthetic enzyme reaction of prokaryotes. *J. Biol. Chem.* **275**, 7462–7465 (2000).
9. Bureau, T. E. & Wessler, S. R. *Tourist*: a large family of small inverted repeat elements frequently associated with maize genes. *Plant Cell* **4**, 1283–1294 (1992).
10. Bureau, T. E., Ronald, P. C. & Wessler, S. R. A computer-based systematic survey reveals the predominance of small inverted-repeat elements in wild-type rice genes. *Proc. Natl Acad. Sci. USA* **93**, 8524–8529 (1996).
11. Jurka, J. & Kapitonov, V. V. *PIFs* meets *Tourists* and *Harbingers*: a superfamily reunion. *Proc. Natl Acad. Sci. USA* **98**, 12315–12316 (2001).
12. Feschotte, C., Jiang, N. & Wessler, S. R. Plant transposable elements: where genetics meets genomics. *Nature Rev. Genet.* **3**, 329–341 (2002).
13. Van den Broeck, D. *et al.* Transposon display identifies individual transposable elements in high copy number elements. *Plant J.* **13**, 121–129 (1998).
14. Fukada-Tanaka, S., Inagaki, Y., Yamaguchi, T., Saito, N. & Iida, S. Colour-enhancing protein in blue petals. *Nature* **407**, 581 (2000).

15. Vos, P. *et al.* AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.* **23**, 4407–4414 (1995).
16. Maes, T., De Keukeleire, P. & Gerats, T. Plant tagmology. *Trends Plant Sci.* **4**, 90–96 (1999).
17. Hirochika, H., Sugimoto, K., Otsuki, Y., Tsugawa, H. & Kanda, M. Retrotransposons of rice involved in mutations induced by tissue culture. *Proc. Natl Acad. Sci. USA* **93**, 7783–7788 (1996).
18. Hirochika, H. *Molecular Biology of Rice* (ed. Shimamoto, K.) 43–58 (Springer, New York, 1999).
19. Hirochika, H. Contribution of the *Tos17* retrotransposon to rice functional genomics. *Curr. Opin. Plant Biol.* **4**, 118–122 (2001).
20. Kumar, A. & Hirochika, H. Applications of retrotransposons as genetic tools in plant biology. *Trends Plant Sci.* **6**, 127–134 (2001).
21. Nakazaki, T. & Ikehashi, H. Genomic sequence and polymorphisms of a rice chitinase gene, *Ch4*. *Breed. Sci.* **48**, 371–376 (1998).

Supplementary Information accompanies the paper on Nature's website

(<http://www.nature.com/nature>).

Acknowledgements We thank H.-Y. Hirano for providing information before publication; and H. Yamagata and H. Hirochika for discussions.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.T.

(e-mail: tanisaka@kais.kyoto-u.ac.jp). The *Rurm1*tm and *Rurm1*⁺ sequences have been deposited in the DNA Data Bank of Japan under accession numbers AB072932 and AB072933, respectively.

Intracardiac fluid forces are an essential epigenetic factor for embryonic cardiogenesis

Jay R. Hove[†], Reinhard W. Köster^{†,‡}, Arian S. Forouhar^{*}, Gabriel Acevedo-Bolton^{*}, Scott E. Fraser[‡] & Morteza Gharib^{*}

^{*}Options of Bioengineering and Aeronautics, Division of Engineering & Applied Science; and [‡]Biological Imaging Center, Beckman Institute, and Division of Biology, California Institute of Technology, Pasadena, California 91125, USA
[†]These authors contributed equally to this work.

The pattern of blood flow in the developing heart has long been proposed to play a significant role in cardiac morphogenesis. In response to flow-induced forces, cultured cardiac endothelial cells rearrange their cytoskeletal structure and change their gene expression profiles^{1,2}. To link such *in vitro* data to the intact heart, we performed quantitative *in vivo* analyses of intracardiac flow forces in zebrafish embryos. Using *in vivo* imaging, here we show the presence of high-shear, vortical flow at two key stages in the developing heart, and predict flow-induced forces much greater than might have been expected for micro-scale structures at low Reynolds numbers. To test the relevance of these shear forces *in vivo*, flow was occluded at either the cardiac inflow or outflow tracts, resulting in hearts with an abnormal third chamber, diminished looping and impaired valve formation. The similarity of these defects to those observed in some congenital heart diseases argues for the importance of intracardiac haemodynamics as a key epigenetic factor in embryonic cardiogenesis.

The formation of a functional heart is regulated by the coordinated interplay between a genetic programme, fluid mechanical stimuli, and the inter- and intracellular processes that link them^{3–5}. While the genetics of cardiogenesis are being analysed intensely^{6,7}, studies of the influence of epigenetic factors such as blood flow on heart development have advanced more slowly owing to the difficulty of mapping intracardiac flow *in vivo*. Numerous *in vitro* studies demonstrate that vascular endothelial cells can both sense and transduce biomechanical stimuli such as wall shear stress and transmural pressure caused by pulsatile blood flow⁸. For example,

cultured vascular endothelial cells can respond to frictional stresses as low as 0.2 dyn cm^{-2} (ref. 9). The applicability of such *in vitro* data will remain speculative without a quantitative understanding of the flow conditions within the intact developing vertebrate heart.

Small size, optical clarity and external development have made zebrafish (*Danio rerio*) an ideal model for *in vivo* studies of the cellular and molecular events of cardiogenesis¹⁰. To take advantage of these properties, we used zebrafish to quantify the haemodynamics in the developing vertebrate heart. Two different embryonic stages were chosen for detailed analysis: 4.5 days post-fertilization (d.p.f.), after key morphogenetic events have generated a heart that is very close to the adult configuration; and 37 hours post-fertilization (h.p.f.), when the forming heart is not much more than a primitive contractile tube. We defined the gross cardiac dynamics at 4.5 d.p.f. using high-speed transmitted light microscopy. Vessels draining the yolk sac through the sinus venosus fill the atrium; atrial contraction forces the blood into a single ventricle from which it is pumped through the bulbus arteriosus, and into the ventral aorta (Fig. 1a, b; see also movie 1 in Supplementary Information). High-speed laser-scanning microscopy of the heart in homozygous tie2::GFP transgenic embryos, which express green fluorescent protein (GFP) in the cardiac valves¹¹, shows that both atrio-ventricular (AV) and ventriculo-bulbal (VB) valves are present and functional at 4.5 d.p.f. (see movie 2 in Supplementary Information). Transmitted-light laser scanning microscopy reveals distinct AV and VB valve dynamics. The VB valve appears stiff, opening in a zipper-like manner beginning from the ventricular side during ventricular systole (Fig. 1c, d) as pressure rises at the onset of ventricular contraction (see movies 2 and 3 in Supplementary Information). The VB valve closes at the end of ventricular systole assuring unidirectional blood ejection through the elastic bulbus and into the ventral aorta. In comparison, the AV valve (and AV region) appears more flexible, undergoing a 2.8-fold diameter change during the cycle of systole and diastole (Fig. 1e, f; see also movie 4 in Supplementary Information).

To visualize the blood flow patterns inside the heart chambers, we used a fluorescent dye (BODIPY-ceramide) to label the blood serum for high-speed fluorescent confocal imaging. The flow patterns of unlabelled blood cells were revealed as dark streaks in the bright fluorescence of the serum. As the time to scan an entire frame (0.7 s) is in the range of one cardiac cycle, the overall shape of the moving heart is distorted; both the speed and the direction of flow will influence the shape and length of the dark streaks. Despite these distortions, the local relationship between neighbouring streaks can be assessed in the confocal images because the raster scanning of the laser beam interrogates neighbouring regions within microseconds along the line-scan direction and within milliseconds in the field-scan direction. Individual erythrocytes move rapidly from the atrium to the ventricle during diastole (Fig. 2a, also see movie 5 in Supplementary Information); laminar flow is indicated by the parallel trajectories through the VB valve (Fig. 2b). Streak lengths represent the movement of the erythrocytes during the image acquisition period, and therefore are proportional to the speed of the blood cell movement. The lengths and linearity of the streaks suggest high ejection velocities and a jet-like blood flow through the VB valve.

The enhanced contrast between the blood serum and endothelial cells provided by the BODIPY-ceramide staining permitted us to visualize the blood volume in the beating 4.5 d.p.f. zebrafish ventricle. Assuming an ellipsoidal volume for the ventricle¹², we calculated an ejection fraction of around 60%, a value surprisingly close to that of the adult human left ventricle ($59 \pm 7\%$) (ref. 13). A large volume ejected rapidly suggests that the blood efflux from the narrow VB valve creates significant shear forces on the cardiac endothelial cells. Combining the ejected volume, apparent ventricular and valve dimensions, and the measured time of contraction yields an estimated intracardiac blood flow of nearly 1 cm s^{-1} .

To characterize more precisely the intracardiac flow and fluid forces, we used high-speed cine imaging at 440–1,000 frames s^{-1} and digital particle image velocimetry (DPIV)¹⁴. High framing rates allow the course of small groups of erythrocytes to be followed through the beating heart. DPIV revealed the velocity vector field generated by the fluid motions as vectors indicating the directionality and relative velocity (encoded in different colours) at each instant during the heart cycle. Erythrocytes moving through the AV constriction during diastole reached velocities of about 0.5 cm s^{-1} ; similar velocities were measured on either side of the VB constriction during ventricular systole (Fig. 2c, d; see also movie 6 in Supplementary Information). Superficial tissue masks erythrocyte movement through the VB valve itself by physically obstructing the optical path and by contributing a non-moving tissue component to the average reported for that DPIV interrogation window. As the latter effect is greatest near chamber walls and in small cross-sectional areas, our estimates of velocity through the VB valve are conservative. DPIV analysis requires that the captured three-dimensional (3D) erythrocyte movements be projected onto a two-dimensional (2D) plane. This dimensional collapse underestimates distances travelled by a given cell in 3D space, further suggesting that our calculated intracardiac velocities of around 0.5 cm s^{-1} should

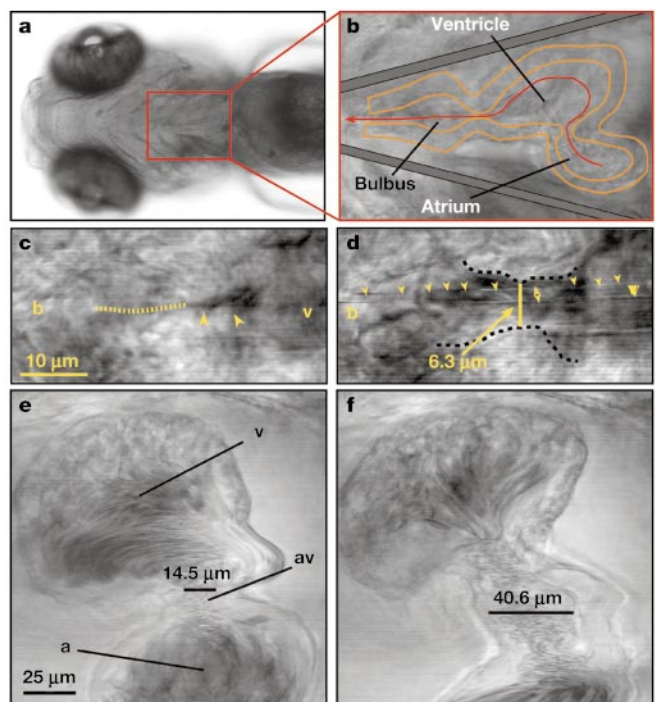


Figure 1 Cardiac dynamics in the zebrafish embryonic heart at 4.5 d.p.f. Pictures display individual characteristic confocal sections from a time series of the embryonic heart beat cycle. At 20 °C the 4.5-d.p.f. zebrafish heart beats at about 2 Hz. **a**, Ventral view of a 4.5-d.p.f. embryo with heart contained in red box. **b**, High magnification of heart with overlaid scheme (in orange) denoting chamber boundaries; for heart beat dynamics see movie 1 in Supplementary Information. Coronal sections through the ventriculo-bulbal (VB) region displaying gating stages of the VB valve. Opening of the valve starts at the ventricular end (yellow arrowheads) and proceeds in a zipper-like fashion (**c**, yellow dashed line indicates the seam of the VB valve) at the beginning of ventricular systole owing to increasing pressure. **d**, At maximal flow gating, diameter ranges from 6 to 10 μm . Note streaks (yellow arrowheads) left by blood cells passing through the valve at high speed. To view *in vivo* valve dynamics see movie 2 and movie 3 in Supplementary Information. Coronal sections taken during ventricular filling (**e**) show that the atrial ventricular (AV) junction is narrow (14.5 μm) and short but enlarges (approximately 2.8-fold) and expands during ventricular systole (**f**) to 40.6 μm ; see movie 4 in Supplementary Information. a, atrium; av, atrio-ventricular junction; b, bulbus arteriosus; v, ventricle.

be taken as lower limits of the real velocities. Flow through the micro-scale cardiac structures of zebrafish embryos is highly viscous (Reynolds number (Re) much less than one) meaning that high-speed flow produces enormous physiological wall shear stresses (with $Re = 0.02$, shear $>75 \text{ dyn cm}^{-2}$) in the VB orifice.

DPIV identifies the circulation patterns of the fluid flow field as a whole (Fig. 2e, f) and reveals the presence of vortices inside the beating embryonic heart. Although vorticity is 3D, vortex circulation is the same in 2D and 3D flow fields, regardless of the 2D plane sectioned (Kelvin's law). During the diastolic phase of the heart cycle, vortices are present behind the AV constriction in the relaxing ventricle (Fig. 2e). This vortical flow pattern (without flow separation), probably caused by the impulsive influx from the AV constriction, is in good agreement with the observed curved trajectories at the AV constriction in BODIPY-ceramide labelled embryos (compare Fig. 2a, e). During systole, DPIV reveals a vortex pair just downstream of the VB orifice, accompanying the blood influx from the ventricle into the bulbus during ventricular contraction (Fig. 2f). This pronounced vortex formation probably has implications for the shape of the bulbus itself, as the size of the bulbus is just sufficient to contain it. The presence of such complex vortical flow in the living zebrafish suggests that *in vitro* studies

should use oscillatory flow regimes rather than simple steady flows to simulate intracardiac flow conditions better *in vivo*.

Such high-shear intracardiac flow during late embryonic stages of cardiac morphogenesis led us to investigate the intracardiac flow patterns in the early embryonic heart (37-h.p.f. zebrafish embryo, Fig. 3a). The dominant morphological feature of the 37-h.p.f. heart is the alignment of the atrium and ventricle in a primitive tube, and the absence of a bulbus (Fig. 3b). Valves will not develop for several hours, consistent with the absence of GFP-labelled valve structures in *tie2::GFP* transgenic embryos; instead, GFP is expressed throughout the endocardium. The 37-h.p.f. heart demonstrates a phasic contraction of the ventricle and atrium at little more than half the rate (1.3 Hz at 20°C) of the 4.5-d.p.f. embryo. Using Nomarski optics we were able to visualize two converging blood streams over the yolk sac being drawn into the atrium (see movie 7 in Supplementary Information). On the basis of individual erythrocyte velocities ($\sim 0.5 \text{ mm s}^{-1}$) and the vessel dimensions, these streams should exert shear forces in excess of 1 dyn cm^{-2} . Despite its lack of valves, the early heart produced a unidirectional flow with little regurgitation, suggesting that the chamber dynamics themselves may govern the blood flow direction¹⁵. To analyse the chamber dynamics in detail, we made use of a transgenic

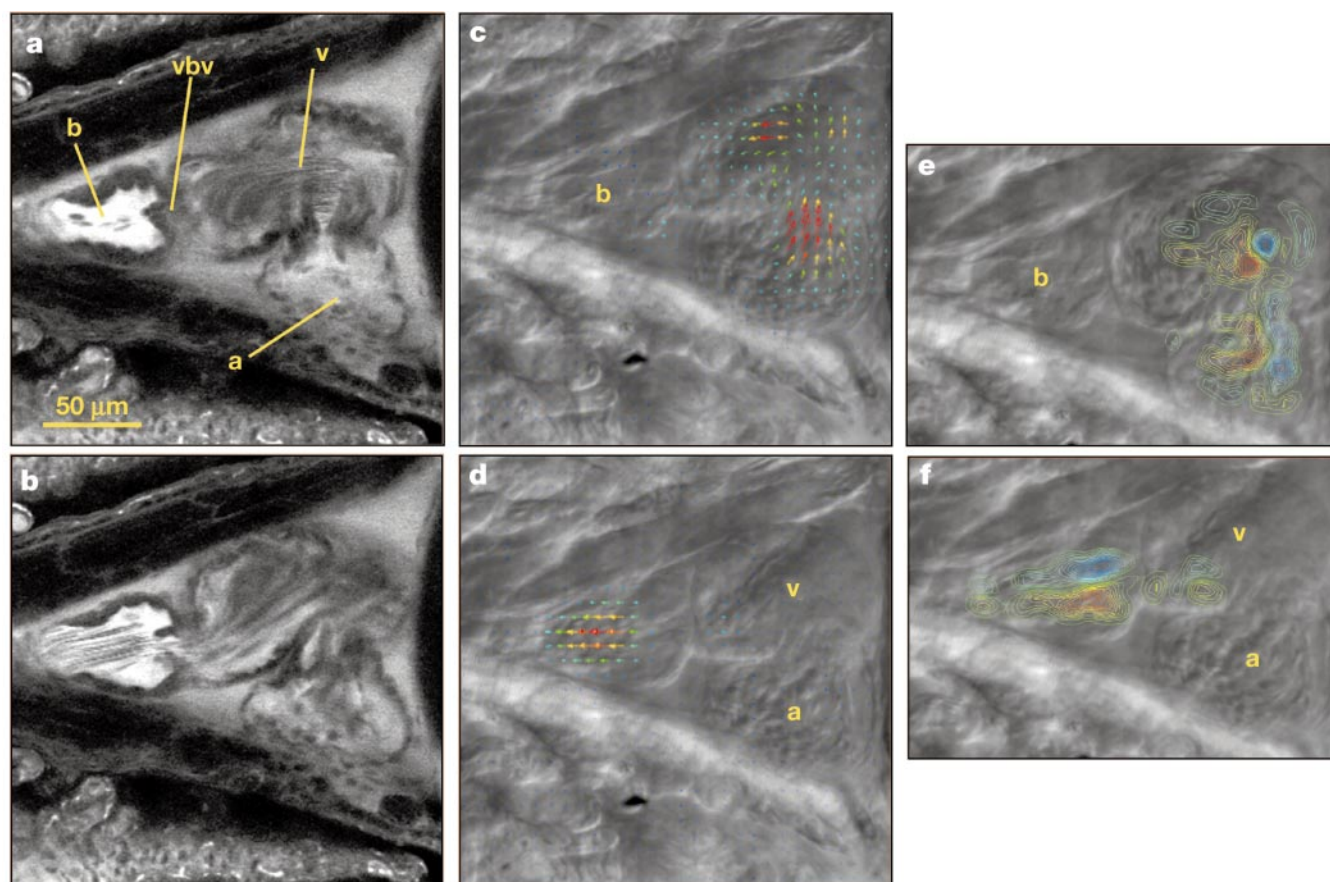


Figure 2 High-velocity, high-shear conditions generated in the 4.5-d.p.f. embryonic zebrafish heart. Pictures display individual characteristic confocal sections from a single time series of BODIPY-ceramide stained embryos. **a**, Atrial systole and ventricular filling. **b**, Ventricular systole and atrial diastole (leading to the refilling of the atrium). Note the laminar ejection pattern generated by forceful microjets from the ventricle through the open VB valve and through the aorta (see also movie 5 in Supplementary Information). **c**, Overlay of digital particle image velocimetry (DPIV) velocity field from real-time, high-speed imaging. Vectors indicate velocity from highest (red), high (yellow), low (green) to lowest (blue). **d**, DPIV velocity field overlaid on bulbus arteriosus to visualize high-velocity transaortic jet. DPIV analysis quantitatively represents blood flow patterns and vector

overlays reveal the fluid jet during diastole (**c**) and systole (**d**). **e**, **f**, Calculated vorticity field from real-time, high-speed imaging at ventricular diastole (**e**) and ventricular systole (**f**). Note the vortical flow behind the AV constriction during ventricular filling (**e**), and the significant vortex pair in the bulbus during ventricular systole (**f**). Contour colours are indicative of strength and direction of rotation as in **c**. Inset shows the intra-bulbar vorticity, computed from velocity vector field, defining the jet. See movie 6 in Supplementary Information for dynamic changes of intracardiac vorticity field and velocity in the beating embryonic heart *in vivo*. a, atrium; b, bulbus arteriosus; v, ventricle; vbv, ventriculo-bulbar valve.

strain expressing GFP under control of the *gata1*-enhancer/promoter¹⁶, which exhibits strong GFP fluorescence in the endocardium, allowing visualization of the heart chamber walls (Fig 3c–e). A nearly complete collapse of a portion of the chambers during systole (Fig. 3d, e; see also movie 8 in Supplementary Information) severely constricts the luminal cross-section, minimizing backflow. The notion that the early embryonic heart may well act as both a valve and pump is intriguing, and current investigations support this hypothesis (J.R.H., G.A.-B. & M.G., manuscript in preparation).

To determine the ejection fraction in the 37-h.p.f. zebrafish heart, BODIPY-ceramide serum labelling was used to visualize the diameter changes of the cylindrical ventricle between maximal systolic contraction and diastolic expansion (Fig. 3f, g) in different transverse planes. The resulting data indicate an ejection fraction of around 60%. The calculated systolic ejection velocity of approximately 1.5 mm s^{-1} is consistent with the maximum blood velocity

measured directly in the AV tract (0.9 mm s^{-1}); these measures predict shear stresses ranging from $2.5\text{--}10 \text{ dyn cm}^{-2}$ in this heart with no valves.

Our *in vivo* data demonstrate that the embryonic zebrafish heart with valves (4.5 d.p.f.) functions as a high-speed, dynamic pump operating on a size scale ($195 \mu\text{m}$ long $\times$ $150 \mu\text{m}$ wide) comparable to a coarse human hair. Calculations, based upon the observed velocities, predict wall shear stresses ranging from 2.5 dyn cm^{-2} in the 37-h.p.f. heart to 76 dyn cm^{-2} in the 4.5-d.p.f. heart. Shear forces less than 1 dyn cm^{-2} are detectable by endothelial cells *in vitro*, resulting in up- or downregulation of gene expression⁹. Larger shear forces, on the order of $8\text{--}15 \text{ dyn cm}^{-2}$, are known to cause cytoskeletal rearrangement¹⁷. Thus, both the early (37 h.p.f.) and the later embryonic hearts (4.5 d.p.f.) exhibit shear forces significantly above the threshold of detectability by the cardiac endothelial cells deduced from *in vitro* data. This strongly suggests that shear forces play some role in the regulation of cardiac morphogenesis during embryogenesis.

To test the possible influences of flow through the embryonic heart on its development, we interfered with blood flow at the primitive heart stage (37 h.p.f.) and analysed the effects on cardiogenesis at the advanced heart stage (4.5 d.p.f.). Beads ($\sim 50 \mu\text{m}$ diameter) were implanted into homozygous transgenic *tie2::GFP* embryos either in front of the sinus venosus to block blood influx into the atrium (Fig. 4b); or in the back of the ventricle to block blood efflux from the heart into the aorta (Fig. 4c). Continuous successful blocking of blood flow (assayed 20 h after bead implantation) resulted in an accumulation of erythrocytes in front of the atrium (Fig. 4e, arrow) or inside the heart chambers respectively (Fig. 4f). In both cases, the embryos showed severe regurgitation of blood inside the heart and reduced blood flows, resulting in dramatically reduced ($\sim$ tenfold) shear forces.

Embryos with impaired cardiac flow demonstrated three dramatic phenotypes. First, their hearts did not form the third chamber, the bulbus. In addition, they lacked heart looping, the normal process resulting in the repositioning of the ventricle and atrium from a cephalo-caudal into a side-by-side arrangement. Finally, the walls of the inflow and outflow tracts collapsed and fused, beginning at 3 d.p.f. (Fig. 4h, i). This latter phenotype is reminiscent of the zebrafish *jekyll* mutant, which demonstrates abnormal blood flow due to a missing AV valve (E. Walsh, D. Stainier, personal communication). The similarity of the defects resulting from disrupting either inflow or outflow suggests that it was not the changes in transmural pressures in the cardiovascular system that is responsible, as the two different blockades should decrease (blocked inflow) or increase (blocked outflow) the pressure. Instead, the decreased shear forces, the common feature of the two treatments, seem most likely to generate the developmental phenotype.

Control experiments and other considerations support the conclusion that altered shear forces result in the observed cardiac defects. First, control embryos, in which a bead was placed close to the sinus without occluding blood flow through the heart (Fig. 4a, d), showed no obvious change in heart morphology or function at 4.5 d.p.f. (Fig. 4g). Additional control embryos, in which inserted beads were removed after one hour, similarly showed no cardiac abnormalities ($n = 5$, data not shown). Thus, cardiac malformations do not result from either the surgical procedure or the continued presence of a bead. In chick embryos, proper heart looping has been reported after the heart beat and thus haemodynamic forces have been blocked; however, the results were scored after only five hours without a heart beat¹⁸. In contrast, morphological defects attributable to improper looping could be observed in venous clipping experiments that lasted for several days¹⁹. It is unlikely that our observed cardiac abnormalities result from subcritical levels of hypoxia as suggested elsewhere²⁰. Zebrafish embryos do not seem to be sensitive to erythrocyte-mediated oxygen delivery¹⁵, because a mutant specifically lacking erythrocytes

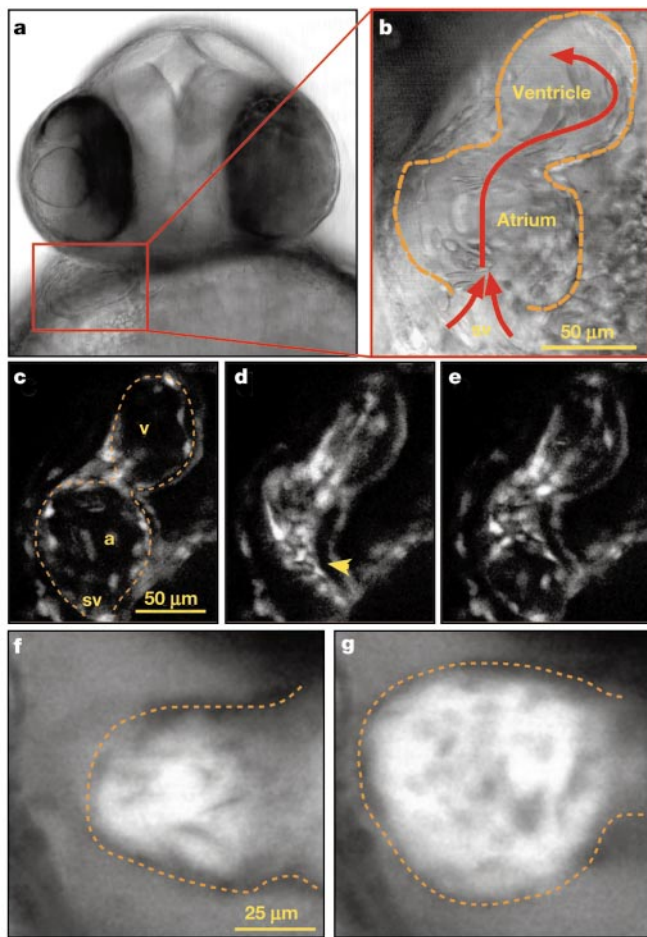


Figure 3 Dynamics of valve-less atrio-ventricular junction in the 37-h.p.f. embryonic zebrafish heart. **a**, Embryo with heart contained in the red box. **b**, High magnification of the heart with overlaid scheme (in orange) denoting chamber boundaries. Red arrows indicate mean flow direction. Note streak-like imprints left by moving blood cells during scan indicating their flow direction and their velocity. For influx of blood over the yolk into the atrium through the developing sinus venosus see movie 8 in Supplementary Information. **c–e**, Confocal coronal sections of different stages of a heart beat cycle of a 37-h.p.f. homozygous *gata1::GFP* embryo, the endocardium and blood cells are labelled by GFP fluorescence. **c**, Diastole of atrium and ventricle with chamber outlines indicated by dashed orange lines. **d**, Maximal atrial systole; note the large reduction in luminal volume by almost complete collapse of the endocardium (yellow arrowhead). **e**, Beginning of atrial diastole and ventricular systole. See Supplementary Information for movie 8. **f**, Ventricle at systole and **g**, diastole for ejection fraction calculations. To counterstain blood serum for better volume measurements embryos were incubated in BODIPY-ceramide. a, atrium; sv, sinus venosus; v, ventricle.

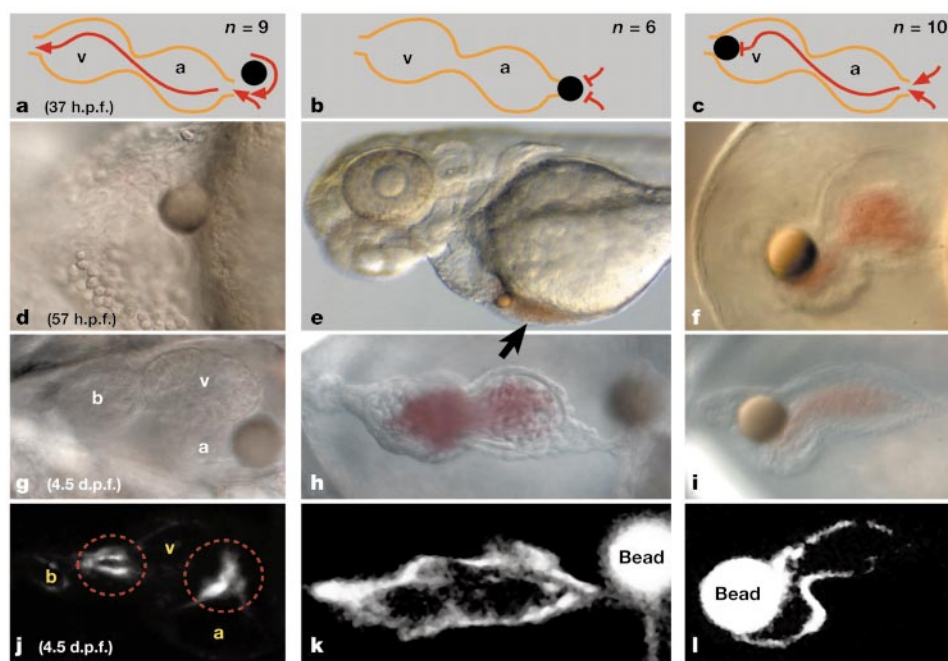


Figure 4 Impaired blood flow influences cardiogenesis. Glass beads (50 μm) were inserted into homozygous transgenic *tie2::GFP* embryos at the primitive heart stage (37 h.p.f.) close to the sinus venosus but not blocking flow (**a**, $n = 9$), in front of the sinus to block blood influx (**b**, $n = 6$) or into the outflow tract to block blood efflux (**c**, $n = 10$). Successful blockage was checked 20 h after surgery (**d–f**, note accumulated blood erythrocytes on the yolk in front of the bead while the atrium contains no erythrocytes (**e**, arrowhead). Operated specimens were re-examined at 4.5 d.p.f. **g**. Cardiac chamber development appears unaffected by non-occlusive implantations and both the AV and VB valves form properly (**j**, red dashed circles, see also movie 2 in Supplementary

Information). In contrast, cardiogenesis in embryos with blocked blood flow is severely disrupted independently of blockage location. Bulbus formation is greatly reduced or fails, heart looping does not occur; and the lateral walls of the sinus and outflow tract collapse and fuse, sealing off the heart (**h**, **i**). In addition, AV and VB valves do not form, initial endocardial GFP expression does not concentrate in the forming valves consistent with the peristaltic, primitive-heart-like, contraction of the affected hearts (**k**, **l**; movies 9 and 10 in Supplementary Information). Except for a yolk-sac oedema, all other organs and tissues in operated-on embryos developed properly and exhibited normal neurological responses such as escape reflex and eye movements. a, atrium; b, bulbus; v, ventricle.

for the first five embryonic days displays no vascular defect and can be raised to adulthood²¹. The defects are unlikely to result from reduced transport of nutrients or other biomolecules, as both our operated-on embryos and mutants lacking blood circulation develop properly over the first few days²². For example, *silent heart* mutant embryos establish gut peristalsis and are able to hatch and swim²³. Swimming is a complicated behaviour, initiated at about 5 d.p.f., requiring the orchestration of a functional nervous system with properly organized musculature. Most cardiac zebrafish mutants die at around day 7, indicating that proper heart function and nutrient delivery is essential for cardiac morphology only significantly later than 4.5 d.p.f. (ref. 22). The normal development of our control embryos, combined with these other findings, suggest that our observed cardiac phenotypes are likely to result from altered intracardiac haemodynamics rather than trauma to the vascular wall and endothelium, hypoxia, or failure of nutrient delivery.

The intracardiac regurgitation of blood observed in our flow-impaired zebrafish embryos points to a failure in valve formation, and the pattern of *tie2::GFP* expression confirms abnormal valve formation under conditions of altered haemodynamics. In contrast to normal *tie2::GFP* embryos, in which the initial GFP expression throughout the endocardium resolves to the forming valves (AV and VB, Fig. 4j, red dashed circles; see movie 2 in Supplementary Information), specimens with impaired flow showed no signs of forming valves. Both blocked-inflow and blocked-outflow embryos displayed peristaltic heart contraction at 4.5 d.p.f. with weak GFP fluorescence throughout the endocardium (Fig. 4k, l; see also movies 9 and 10 in Supplementary Information). Thus, direct interference with intracardiac haemodynamics has profound effects on cardiac valve formation, as well as chamber differentiation and organ morphology during embryogenesis. Approximately 40% of

congenital heart defects involve a valve abnormality as a contributing component (D. Sahn, personal communication). Our defects in hearts with no genetic lesion suggest that a critical role is played by blood flow-induced forces during normal heart development and suggests that altered haemodynamics may contribute to the cardiac phenotype in some cardiac mutants and perhaps also birth defects.

Through the application of these flow visualization technologies²⁴, we have quantified embryonic intracardiac velocities and shear forces directly *in vivo*. Surprisingly, the forces reach levels significantly above cellular sensitivity even in the primitive heart of the early embryo. Interference with these intracardiac flow forces results in severe cardiac malformations. Thus, we conclude that the physiology and development of the embryonic heart is linked by haemodynamics, consistent with recent observations on cardiac pathologies. There is increasing evidence that cardiovascular disease is rooted in processes initiated in the developing embryo and fetus²⁵. The results presented here extend this proposal, suggesting that the same responses to shear forces that shape the developing heart may contribute to later abnormalities and disease processes. Our findings underscore the importance of examining the interplay between genetics and epigenetic factors such as fluid forces in analysing the pathogenesis of embryonic vascular defects and cardiovascular disease^{26,27}. □

Methods

Maintenance of fish

Embryos were raised in 30% Danieau solution. Raising, maintaining and spawning of adult zebrafish were performed as described^{28,29}. Adding phenylthiourea (PTU) in a concentration of 0.15 mM to the embryonic raising medium at the 10-somite stage (about 14 h.p.f.) blocked pigmentation. Dechorionated embryos were anaesthetized with 0.05% tricaine, embedded in 1.2% ultralow gelling agarose (Sigma) and mounted on a coverslip.

In vivo imaging

Confocal imaging was performed at 20 °C using a Zeiss LSM400 laser-scanning microscope. Imaging at this temperature slowed the beat frequency down to allow better imaging but did not change haemodynamics. For flow pattern and ejection fraction analysis embryos were soaked in 0.001% BODIPYFL C₅-ceramide (Molecular Probes) as a contrast agent overnight before imaging. Times for continuous high-speed scanning to image flow dynamics were chosen such that individual pictures were out-of-phase compared to the heart beat cycle to capture subsequent stages of diastole and systole. Thus movies obtained from confocal imaging represent composites. Confocal images were analysed using LSM software (Zeiss), Object-Image 2.05, Photoshop 5.02 (Adobe) and Freehand 8 (Macromedia).

Images for DPIV analysis were acquired at 256 × 256 pixels in real time using a Dalsa CA-D6-0256 charge-coupled device (CCD) camera, and processed using interrogation windows of 16 × 16 pixels with an 8 × 8 pixel overlap (50%). DPIV analysis was performed on images acquired at 20, 25 and 28 °C. The intracardiac blood flow patterns appeared to be temperature invariant across this range although flow velocity did decrease slightly with lower temperature. For this reason the data sets taken at 20 °C were used for final DPIV analysis as they delivered a higher number of frames per heart beat cycle.

Ejection fraction calculation

The volume at the two extremes of the heart cycle was calculated by taking half of the projection through the mid-plane of the heart in two different frames corresponding to the maximum volume and minimum volume. The heart was assumed to be symmetrical across this line. For each frame a curve was fitted, $f(x)$, to the contour of heart wall in this half-section and rotated about the symmetry line. Then the formula for volume of revolution was applied:

$$V = \pi \int_a^b [f(x)]^2 dx \quad (1)$$

Change in volume was calculated by taking the difference in volume between the two frames, ΔV .

Velocity and shear stress calculation

Without time for a parabolic flow profile to evolve, we assumed a linear velocity profile in the regions of interest, with maximum velocity at the centre, and zero velocity at the walls. Velocity was expressed by the equation:

$$u(y) = \frac{U}{a}y \quad (2)$$

where U was the centreline velocity, a was the half-width of the region of interest (that is, the radius) and y was the distance from wall, reaching a maximum value of a at the centre. Shear stress (τ) was defined as the derivative of velocity multiplied by μ , the dynamic viscosity of the fluid, such that:

$$\tau = \mu \frac{\partial u(y)}{\partial y} \quad (3)$$

To convert into dyn cm⁻² the result was multiplied by a factor of ten. Given the small size of the vessel under examination, μ was assumed to be constant at 5×10^{-3} N s m⁻², roughly five times that of water.

Cardiac occlusion surgery

Embryos (tie2::GFP; 37 h.p.f.) were anaesthetized in 0.05% tricaine and embedded in low-melting agarose. The epithelium in front of the heart was cut open using micromanipulation needles; beads (BioRad AG1X8 Cl⁻) were inserted into the bloodstream and positioned with tungsten needles. The size of the bead (~50 µm) was chosen to fit into the space between the yolk and the overlying skin, close to the sinus venosus. Beads were kept in place by the slight pressure of these two tissues but could be repositioned as necessary.

Received 9 August; accepted 31 October 2002; doi:10.1038/nature01282.

- Topper, J. N. & Gimbrone, M. A. Blood flow and vascular gene expression: fluid shear stress as a modulator of endothelial phenotype. *Molec. Med. Today* **5**, 40–46 (1999).
- Davies, P. F. Flow-mediated endothelial mechanotransduction. *Physiol. Rev.* **75**, 519–560 (1995).
- Nerem, R. M., Harrison, D. G., Taylor, W. R. & Alexander, R. W. Hemodynamics and vascular endothelial biology. *J. Cardiovasc. Pharmacol.* **21**, S6–S10 (1993).
- Takahashi, M., Ishida, T., Traub, O., Corson, M. A. & Berk, B. C. Mechanotransduction in endothelial cells: Temporal signaling events in response to shear stress. *J. Vasc. Res.* **34**, 212–219 (1997).
- Taber, L. A. Mechanical aspects of cardiac development. *Prog. Biophys. Mol. Biol.* **69**, 237–255 (1998).
- Chen, J. N. & Fishman, M. C. Genetics of heart development. *Trends Genet.* **16**, 383–388 (2000).
- Srivastava, D. & Olson, E. N. A genetic blueprint for cardiac development. *Nature* **407**, 221–226 (2000).
- Helminger, G., Geiger, R. V., Schreck, S. & Nerem, R. M. Effects of pulsatile flow on cultured vascular endothelial-cell morphology. *J. Biomech. Eng. Trans. ASME* **113**, 123–131 (1991).
- Olesen, S. P., Clapham, D. E. & Davies, P. F. Hemodynamic shear-stress activates a K⁺ current in vascular endothelial cells. *Nature* **331**, 168–170 (1988).
- Stainier, D. Y. R. & Fishman, M. C. The zebrafish as a model system to study cardiovascular development. *Trends Cardiovasc. Med.* **4**, 207–212 (1994).
- Walsh, E. C. & Stainier, D. Y. R. UDP-glucose dehydrogenase required for cardiac valve formation in zebrafish. *Science* **293**, 1670–1673 (2001).
- Hou, P.-C. L. & Burggren, W. W. Cardiac output and peripheral resistance during larval development in the anuran amphibian *Xenopus laevis*. *Am. J. Physiol. Regul. Integr. Compar. Physiol.* **38**, R1126–R1132 (1995).
- Mankad, R. et al. Regional myocardial strain before and after mitral valve repair for severe mitral

regurgitation. *J. Cardiovasc. Magn. Res.* **3**, 257–266 (2001).

- Willert, C. E. & Gharib, M. Digital particle image velocimetry. *Exp. Fluids* **10**, 181–193 (1991).
- Pelster, B. & Burggren, W. W. Disruption of hemoglobin oxygen transport does not impact oxygen-dependent physiological processes in developing embryos of zebra fish (*Danio rerio*). *Circ. Res.* **79**, 358–362 (1996).
- Long, Q. M. et al. GATA-1 expression pattern can be recapitulated in living transgenic zebrafish using GFP reporter gene. *Development* **124**, 4105–4111 (1997).
- Davies, P. F., Remuzzi, A., Gordon, E. J., Dewey, C. F. & Gimbrone, M. A. Turbulent fluid shear-stress induces vascular endothelial-cell turnover *in vitro*. *Proc. Natl Acad. Sci. USA* **83**, 2114–2117 (1986).
- Manasek, F. J. & Monroe, R. G. Early cardiac morphogenesis is independent of function. *Dev. Biol.* **27**, 584–588 (1972).
- Hogers, B., DeRuiter, M. C., GittenbergerdeGroot, A. C. & Poelmam, R. E. Unilateral vitelline vein ligation alters intracardiac blood flow patterns and morphogenesis in the chick embryo. *Circ. Res.* **80**, 473–481 (1997).
- Icardo, J. M. Developmental biology of the vertebrate heart. *J. Exp. Zool.* **275**, 144–161 (1996).
- Liao, E. C. et al. Non-cell autonomous requirement for the *bloodless* gene in primitive hematopoiesis of zebrafish. *Development* **129**, 649–659 (2002).
- Stainier, D. Y. R. et al. Mutations affecting the formation and function of the cardiovascular system in the zebrafish embryo. *Development* **123**, 285–292 (1996).
- Sehnert, A. J. et al. Cardiac troponin T is essential in sarcomere assembly and cardiac contractility. *Nature Genet.* **31**, 106–110 (2002).
- Meinhart, C. D., Wereley, S. T. & Santiago, J. G. PIV measurements of a microchannel flow. *Exp. Fluids* **27**, 414–419 (1999).
- O'Brien, S. P. M., Wheeler, T. & Barker, D. J. P. *Fetal Programming Influences on Development and Disease in Later Life* (RCOG, London, 1999).
- Di Stefano, I., Koopmans, D. R. & Langille, B. L. Modulation of arterial growth of the rabbit carotid artery associated with experimental elevation of blood flow. *J. Vasc. Res.* **35**, 1–7 (1998).
- Langille, B. L. *Flow-dependent Regulation of Vascular Function* (eds Bevan, J. A., Kaley, G. & Rubanyi, G.) 277–299 (Oxford Press, New York, 1995).
- Kimmel, C. B., Ballard, W. W., Kimmel, S. R., Ullmann, B. & Schilling, T. F. Stages of embryonic development of the zebrafish. *Dev. Dyn.* **203**, 253–310 (1995).
- Westerfield, M. *The Zebrafish Book* (Univ. Oregon Press, Eugene, 1995).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements We thank S. Lin for the stable transgenic *gata1::GFP* zebrafish strain, E. Walsh, E. Ober, B. Jungblut and D.Y.R. Stainier for the transgenic *tie2::GFP* strain and comments on the manuscript and M. Bak and Y. Gong for discussions. This work was supported by the American Heart Association (J.R.H.) and the Human Frontier Science Program (R.W.K.), the NIH (S.E.F.), the Beckman Institute at Caltech and the Powell Foundation. J.R.H. worked on flow pattern analysis and calculations. R.W.K. worked on imaging and embryonic manipulation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.R.H. (e-mail: jhove@caltech.edu) or R.W.K. (e-mail: rkoestel@egg.caltech.edu).

Rhythmic histone acetylation underlies transcription in the mammalian circadian clock

Jean-Pierre Etchegaray*, Choogon Lee*, Paul A. Wade† & Steven M. Reppert*

* Department of Neurobiology, University of Massachusetts Medical School, LRB-728, 364 Plantation Street, Worcester, Massachusetts 01605, USA
† Department of Pathology, Emory University, Whitehead Research Building, Room 142, 615 Michael Street, Atlanta, Georgia 30322, USA

In the mouse circadian clock, a transcriptional feedback loop is at the centre of the clockwork mechanism. Clock and Bmal1 are essential transcription factors that drive the expression of three period genes (*Per1–3*) and two cryptochrome genes (*Cry1* and *Cry2*)^{1–5}. The Cry proteins feedback to inhibit Clock/Bmal1-mediated transcription by a mechanism that does not alter Clock/Bmal1 binding to DNA⁶. Here we show that transcriptional regulation of the core clock mechanism in mouse liver is accompanied by rhythms in H3 histone acetylation, and that H3

acetylation is a potential target of the inhibitory action of Cry. The promoter regions of the *Per1*, *Per2* and *Cry1* genes exhibit circadian rhythms in H3 acetylation and RNA polymerase II binding that are synchronous with the corresponding steady-state messenger RNA rhythms. The histone acetyltransferase p300 precipitates together with Clock *in vivo* in a time-dependent manner. Moreover, the Cry proteins inhibit a p300-induced increase in Clock/Bmal1-mediated transcription. The delayed timing of the *Cry1* mRNA rhythm, relative to the *Per* rhythms, is due to the coordinated activities of Rev-Erb α and Clock/Bmal1, and defines a new mechanism for circadian phase control.

A 'master' clock in the hypothalamic suprachiasmatic nuclei (SCN) orchestrates circadian rhythms in mammals⁷. The SCN are entrained (synchronized) to the 24-h day by the daily light/dark cycle acting through retina-to-SCN neural pathways. The SCN clock uses neural, hormonal and behavioural outputs to entrain 'slave' oscillators in other brain areas and in peripheral tissues. These slave oscillators, in turn, generate local rhythms in physiological processes that control tissue-specific functions^{8–10}. Genetic studies have shown that analysis of clock gene function in slave oscillators is a valid approach for defining molecular and biochemical principles important for the central clock mechanism^{6,11}.

As changes in chromatin structure are critical for the regulation of transcription in other systems^{12–14}, we hypothesized that the Clock/Bmal1-mediated rhythmic drive in *Per1* transcription is regulated by circadian changes in chromatin structure. For these studies, we exploited the liver oscillator using formaldehyde-cross-linked chromatin immunoprecipitation (X-ChIP) and semi-quantitative polymerase chain reaction (PCR) analysis of the *Per1* promoter in the chromatin complexes. We examined the binding of

RNA polymerase II (pol II) to the *Per1* promoter to assay for dynamic changes in coactivator assembly necessary for transcriptional activity. We also studied the degree of acetylation of H3 and H4 histones in the promoter to assay for rhythmic changes in chromatin structure.

In contrast to the lack of a substantial oscillation in Clock/Bmal1 binding to *Per1* E boxes⁶, the promoter exhibited robust circadian rhythms in pol II binding and H3 acetylation (Fig. 1a). The acetylation rhythm appeared specific for H3, as no circadian rhythm was found for the acetylation of H4 (data not shown). The pol II binding and H3 acetylation rhythms were synchronous with each other and with the rhythm in *Per1* mRNA levels, with values peaking at around circadian time (CT) 9 (Fig. 1a). Similarly phased circadian rhythms in H3 acetylation and pol II binding were also found for *Per2* (Fig. 1b). The decline in *Per2* mRNA levels is slower, compared with *Per1*, probably reflecting differences in mRNA stability. These results suggest that the rhythmic transcription of *Per1* and *Per2* is associated with circadian rhythms in chromatin remodelling and in transcriptional coactivator recruitment and assembly.

The rhythms in chromatin structure and coactivator assembly at the *Per* promoters provide a potential mechanistic explanation for the ability of the Cry complexes to negatively regulate Clock/Bmal1-mediated transcription. It is possible that the binding of negative regulators through direct Cry–Clock–Bmal1 interactions disrupts the histone acetyltransferase (HAT)-containing transcriptional coactivator complex, which, in turn, disrupts ongoing histone acetylation. This possibility was explored by evaluating time-dependent associations of Clock with p300, Cbp, Gcn5 and Pcaf in liver nuclear extracts. These four proteins have intrinsic HAT activity and are often components of the coactivator transcriptional complex associated with H3 acetylation^{12–14}. In fact, p300 and Cbp can bind directly to Bmal1 in yeast two-hybrid studies and

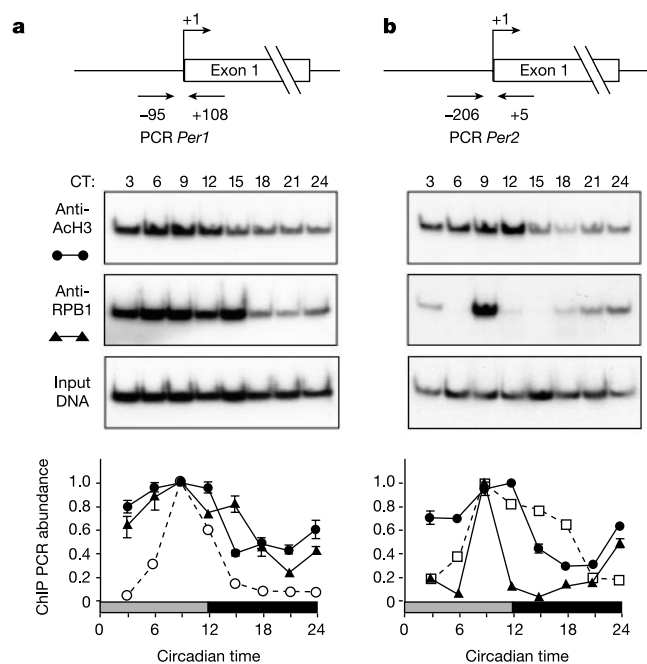


Figure 1 The *Per1* and *Per2* genes exhibit synchronous rhythms in H3 acetylation, pol II binding, and mRNA levels. **a**, *Per1* promoter rhythms. Representative Southern blots of H3 acetylation (anti-Ach3), pol II binding (anti-RPB1), and input DNA are shown. The graph shows relative PCR values; values are mean $\pm$ s.e.m. ($n = 6$ blots). The data are plotted as percentages relative to the highest value (1.0). The dashed line indicates steady-state *Per1* mRNA levels⁶. **b**, *Per2* promoter rhythms. Representative Southern blots are shown. Values are mean $\pm$ s.e.m. ($n = 4$ blots). The dashed line indicates steady-state *Per2* mRNA levels⁶. Horizontal bars at bottom represents the light (grey)/dark (black) cycle prior to placement in constant darkness. CT, circadian time.

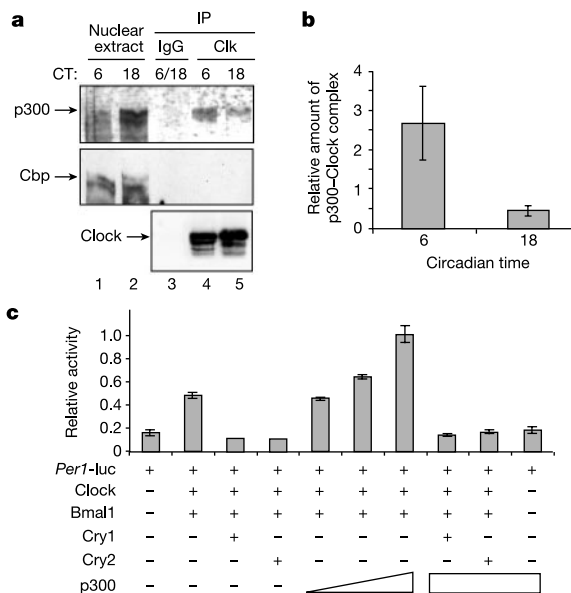


Figure 2 HAT proteins in the clockwork. **a**, p300–Clock associations *in vivo*. Nuclear extracts were immunoprecipitated (IP) with an antibody against Clock or with IgG. **b**, Quantification of relative p300–Clock associations. Values are mean $\pm$ s.e.m. ($n = 3$ sets of nuclei). **c**, Cry inhibits a p300-induced increase in transcription. The reporter contained the *Per1* promoter⁵. Presence (+) or absence (–) of reporter (25 ng) and expression plasmids (2.5 ng of *Clock* and *Bmal1*; 100 ng of *Cry1*; 65, 85 and 100 ng p300) is indicated. Activity is relative to the highest response (1.0). Values are mean $\pm$ s.e.m. ($n = 3$).

accentuate Clock/Bmal1-mediated transcription in cell culture¹⁵.

We found that p300 precipitated together with Clock in liver nuclei and that the amount of p300 associated with Clock was fivefold greater at CT 6 compared with CT 18 ($P < 0.03$, Student's t -test) (Fig. 2a, b). Cbp was detected in whole nuclear extracts, but not in the Clock immune complexes at either circadian time (Fig. 2a). We were unable to detect Gcn5 or Pcaf in whole nuclear extracts at either circadian phase (data not shown). These results indicate that p300 is a member of the Clock–Bmal1–coactivator complex. The night-time decrease in Clock–Bmal1–p300 interactions (and, by inference, HAT activity) may contribute to the decrease in H3 acetylation and transcriptional activity.

We next investigated whether the Cry proteins have a function in the disruption of the Clock–Bmal1–coactivator complex. Luciferase reporter gene assays using the *Per1* promoter demonstrated that the Cry proteins can inhibit a p300-induced increase in Clock/Bmal1-mediated transcription *in vitro* (Fig. 2c). Clock and Bmal1 together caused an increase in transcriptional activity through their previously described actions on three CACGTG E boxes in the *Per1* promoter^{3,4}. p300 amplified the Clock/Bmal1-mediated increase in transcriptional activity in a dose-dependent manner, whereas the addition of p300 without Clock and Bmal1 did not increase basal transcription (Fig. 2c). The Cry proteins inhibited the p300-induced increase in transcriptional activity by over 80%.

At the time that Cry–Clock–Bmal1 interactions seem to decrease HAT activity, the Cry-containing negative regulatory complex might also induce the involvement of a histone deacetylase (HDAC), as occurs in other systems^{12–14}. We found, however, moderate HDAC activity associated with the Clock-containing complexes at both CT 6 and CT 18, and the activity was actually higher at CT 6 (without Cry association) compared with CT 18 (the time of peak Cry–Clock associations) (Supplementary Fig. 1). These results show that the Cry-containing complex does not augment the HDAC activity already in the Clock complex. Instead, the balance between acetylation and deacetylation of H3 seems to be regulated

primarily by the regulation of HAT activity, secondary to rhythmic, Cry-mediated disruption of the Clock–Bmal1–coactivator complex. The combined results support a model in which the Cry proteins disrupt the Clock-associated transcriptional complex, reducing HAT activity and altering chromatin structure to reduce Clock/Bmal1 transcriptional activation, resulting in negative feedback.

Although our *in vivo* studies suggest that the rhythms in chromatin structure are cell autonomous, it is possible that the synchronous *Per1* and *Per2* promoter rhythms are driven by circadian changes in circulating factors. We therefore evaluated the transcriptional regulation of another clock gene, *Cry1*, whose mRNA peak in liver is delayed by 6–9 h from that found for *Per1* and *Per2* (ref. 6). The *Cry1* RNA rhythm is under the control of Clock, as it is severely blunted in both the SCN and slave oscillators of *Clock/Clock* mutant mice⁵. Furthermore, the *Cry1* 5' flanking region contains a CACGTG E box at positions –270 to –265, making the E box a candidate site through which Clock–Bmal1 heterodimers may activate *Cry1* transcription.

We used DNaseI hypersensitivity to map transcriptionally active hypersensitive sites in an 8-kilobase (kb) stretch of the *Cry1* gene that encompasses 1.5 kb of the promoter region and 6.5 kb downstream of the transcription start site. Southern analysis identified hypersensitive sites just downstream of the CACGTG E box in the promoter, and these sites exhibited a daily rhythm in hypersensitivity, the timing of which correlated with the mRNA rhythm (data not shown). Ligation-mediated (LM)-PCR identified the rhythmic presence of two hypersensitive sites at –253 and –247 (Fig. 3a–c). The rhythm of these two sites also correlated with the mRNA abundance rhythm⁶, with all three peaking at CT 18. The DNaseI Southern blot analysis and LM-PCR results support a functional role of the promoter CACGTG E box for rhythmic transcriptional enhancement of the *Cry1* gene. The hypersensitivity data also suggest that rhythmic changes in chromatin structure accompany the rhythmic changes in *Cry1* transcriptional activity.

To examine directly the ability of Clock–Bmal1 heterodimers to

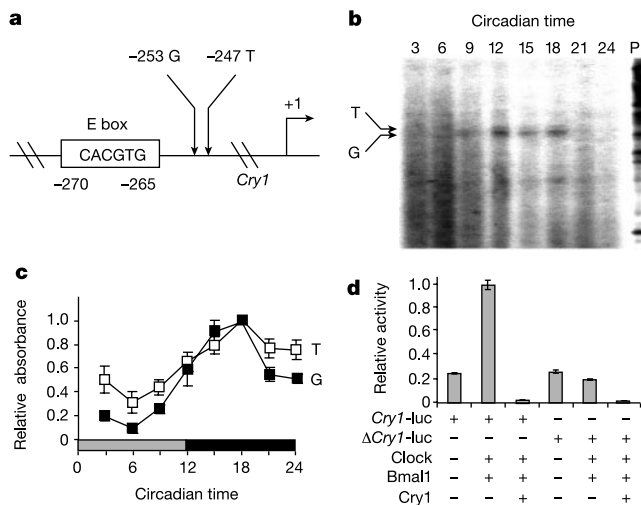


Figure 3 Functional assessment of the *Cry1* promoter. **a**, Location of DNaseI hypersensitive sites (G and T). **b**, Blot of rhythmic occurrence of hypersensitive sites. P, control primer extension. **c**, Quantitative analysis of hypersensitivity rhythms. Values are mean $\pm$ s.e.m. ($n = 3$ experiments). **d**, Clock/Bmal1-mediated activation. Reporters contained either wild-type (*Cry1*-luc) or mutated E box (Δ *Cry1*-luc). Presence (+) or absence (–) of reporter (25 ng) and expression plasmids (250 ng of *Clock*, and *Bmal1*; 100 ng of *Cry1*) is indicated. Activity is relative to response in the presence of Clock and Bmal1 (1.0). Values are mean $\pm$ s.e.m. ($n = 3$).

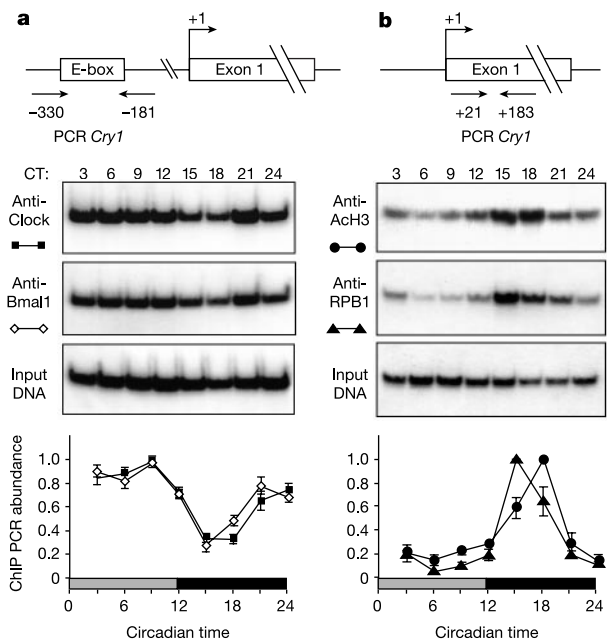


Figure 4 Contrasting rhythms in transcription-factor binding and chromatin remodelling for the *Cry1* gene. **a**, Circadian rhythms in Clock and Bmal1 binding to the promoter E box. Representative Southern blots of the patterns of binding are shown, along with the input DNA control. The graph shows relative PCR values; values are mean $\pm$ s.e.m. ($n = 14$ blots). **b**, Circadian rhythms in H3 acetylation and pol II binding. Representative Southern blots are shown. The graph depicts relative PCR values ($n = 6$ blots).

drive *Cry1* transcription through the E box enhancer, we subcloned a 300-bp fragment of the 5' flanking region containing the endogenous promoter and the CACGTG E box into a promoterless luciferase reporter vector. Using a luciferase reporter gene assay, we found that Clock and Bmal1 together induced a fourfold increase in transcriptional activity through the *Cry1* promoter (Fig. 3d). The Clock/Bmal1-induced activation was dependent on the E box element, because mutation of the CACGTG motif abolished the transcriptional activation by Clock and Bmal1. Clock/Bmal1-induced luciferase activity from the wild-type promoter was blocked by either *Cry1* (Fig. 3d) or *Cry2* (data not shown).

We next examined the *in vivo* binding of Clock and Bmal1 to the *Cry1* E box in liver nuclear extracts using X-ChIP with antibodies against each transcription factor. The results showed that Clock and Bmal1 were both bound to the CACGTG E box in the promoter throughout the circadian cycle, but that the binding patterns exhibited synchronous circadian rhythms of unexpected phase; peak binding occurred at CT 3 to CT 9, with low binding at CT 15 and CT 18 (Fig. 4a). X-ChIP studies of H3 acetylation and pol II binding revealed robust rhythms in exon 1 that were synchronous with the *Cry1* mRNA rhythm, but out of phase with the DNA binding rhythms of Clock and Bmal1 (Fig. 4b). These results show that the binding of Clock and Bmal1 to the promoter, in and of itself, is not sufficient to activate transcription, because binding of these proteins is lowest when the rate of increase in mRNA levels, pol II binding, and H3 acetylation is highest. The phase difference between the *Per* and *Cry* rhythms shows that the chromatin remodelling rhythms are an intrinsic part of the intracellular core clock mechanism.

What explains the different times of transcriptional activation of the *Per* genes compared with the *Cry1* gene, if each is using the same transcription factors bound to CACGTG E boxes? One possibility

is that inhibition of *Cry1* transcription *in vivo* is extended by the rhythmic activity of a repressor. The orphan nuclear receptor Rev-Erb α is the prime candidate for a *Cry1* repressor, because it functions as a transcriptional repressor of *Bmal1* transcription, and its mRNA levels exhibit a circadian rhythm that peaks at CT 6 (refs 16, 17). Moreover, *Cry1* mRNA levels are prematurely elevated (by 8 h) in the livers of Rev-Erb α -deficient mice, whereas the *Per2* mRNA rhythm is unchanged¹⁶. In *Clock/Clock* mutant mice, the blunted *Cry1* mRNA rhythm is also phase advanced by several hours⁵. *Rev-Erb α* mRNA levels are arrhythmic and at low levels in *Clock/Clock* mutant mice¹⁶. These data are consistent with a repressive role of Rev-Erb α on *Cry1* transcription.

We identified three candidate Rev-Erb/ROR binding sites in the *Cry1* gene, one in the 5' flanking region (site I) and two in the first intron (sites II and III; Fig. 5a). All three putative binding sites showed high probability PATSER scores and were a perfect match for six conserved residues in the binding site (Fig. 5a)¹⁷. Electrophoretic mobility shift assays (EMSAs) showed that *in vitro*-translated Rev-Erb α bound specifically to the radiolabelled oligonucleotides surrounding sites I and II, as well as a previously characterized Rev-Erb/ROR site in the *Bmal1* promoter¹⁶ (Fig. 5b, lanes 1, 2 and 5). Similarly, when liver nuclear extracts were incubated with the oligonucleotides, a specific protein-DNA complex was apparent for sites I and II using extracts obtained at CT 6, when Rev-Erb α levels are highest in liver, and not detectable from extracts obtained at CT 18, when protein levels are low (Fig. 5b; data not shown). Of note, the electrophoretic mobility of the specific protein-DNA complexes was retarded by the addition of the anti-Rev-Erb α antibody (Fig. 5c; Supplementary Fig. 2; data not shown).

In a final set of experiments, luciferase reporter gene assays showed that Rev-Erb α can inhibit basal transcription through binding sites I and II in the *Cry1* gene (Fig. 5d). The inhibition

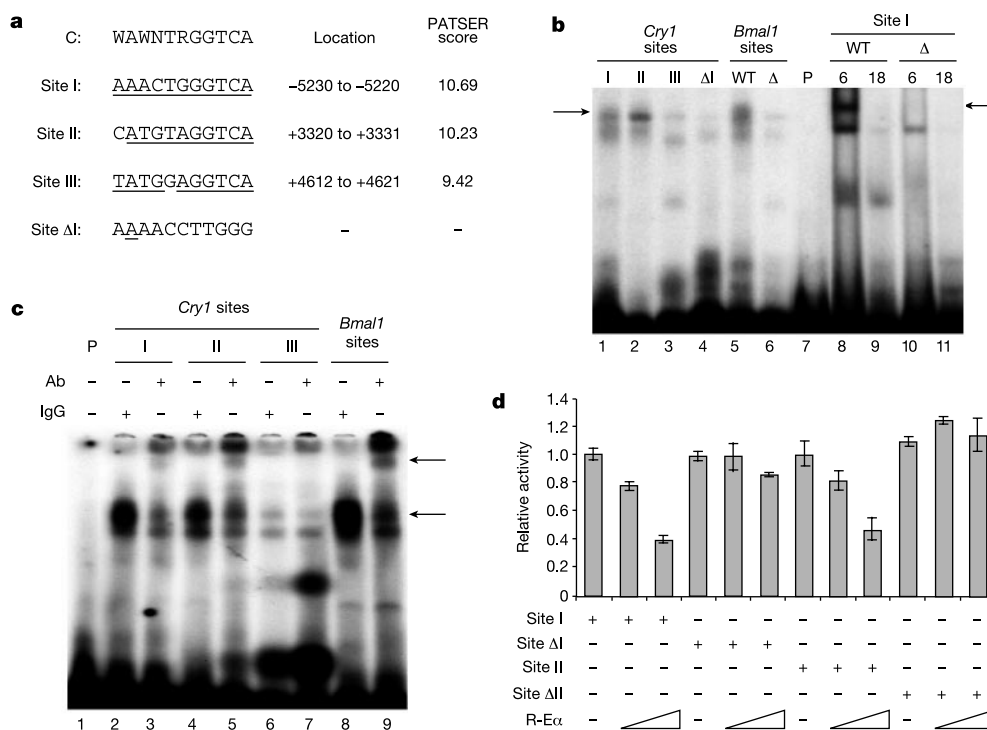


Figure 5 Rev-Erb/ROR binding sites in the *Cry1* gene. **a**, Sequence of putative binding sites. Consensus (C) sequences are underlined. Δ, mutated; PATSER scores, $-\log(10)^{17}$. **b**, Electrophoretic mobility shift assays. Arrows, specific complexes; P, probe alone; 6 and 18, circadian times of collection; WT, wild type. **c**, Super shift EMSAs. Ab, anti-Rev-Erb α antibody. The top and bottom arrows indicate super shift and specific complexes,

respectively. **d**, Rev-Erb α inhibits *Cry1* transcription. Reporters (20 ng each) contained site I, site II, or mutated sites (Δ). Two doses of Rev-Erb α (R-E α ; 10 or 40 ng), were used. Activity is relative to response in the absence of Rev-Erb α (1.0). Values are mean $\pm$ s.e.m. ($n = 3$).

was dose-dependent, and it was eliminated when the binding sites were mutated. Collectively, the EMSAs, luciferase reporter gene studies, and genetic data strongly suggest that Rev-Erb α is a repressor of *Cry1* transcription in liver and that its activity explains the delayed phase of transcriptional enhancement relative to *Per* transcription. Our results define a context-specific transcriptional model for circadian phase control that may be used to regulate the expression of several genes¹⁷. Clock/Bmal1 activation of *Cry1* transcription at CT 18, despite high levels of nuclear Cry proteins, suggests that prior Rev-Erb α -mediated repression alters the sensitivity of the transcriptional activation complex to Cry inhibition.

We propose that a 'histone code' underlies rhythmic transcriptional activity generated by Clock-Bmal1 heterodimers in the circadian clock mechanism. On the basis of studies in other systems^{12–14}, histone acetylation is probably accompanied by parallel changes in other covalent modifications, such as phosphorylation, methylation, ubiquitination and/or ATP-dependent remodelling events. In fact, we have identified circadian rhythms in H3 phosphorylation that parallel those of H3 acetylation for the *Per1* and *Per2* promoters in the liver oscillator (our own unpublished data), and previous studies have shown that light induces H3 phosphorylation in the SCN¹⁸. Deciphering the histone code of the circadian clock mechanism is fundamental to understanding the timekeeping mechanism. □

Methods

Animals

BALB/c mice (Charles River Laboratories) were entrained to a 12-h light/dark cycle for 10 days. At selected times on the first cycle in constant darkness, animals were killed, and livers were dissected and either processed immediately to purify nuclei by sucrose gradient or frozen on dry ice.

Formaldehyde-crosslinked ChIP assays

Six livers were collected every 3 h for the first day in constant darkness. Liver nuclei were isolated and crosslinked with formaldehyde, as described⁶. Each crosslinked sample was sonicated on ice and then incubated with antibodies against acetylated H3 histone at lysines 9 and/or 14 (Upstate Biotechnology), acetylated H4 histone at multiple lysines (residues 5, 8, 12, and/or 16; Upstate Biotechnology), the catalytic subunit of pol II (RPB1; gift from D. Reinberg), Clock⁶ or Bmal1 (ref. 6), as previously described⁶. DNA was isolated from the immunoprecipitates⁶ and subjected to PCR using the following primer pairs: for the promoter region of *Per1*, 5'-CAGATGCCAGGAAGAGATCCTTAGCCA ACC-3' and 5'-GACTAACCCCTAGGATTCGAGCAGGGATCC-3'; for the promoter region of *Per2*, 5'-AGCAGCATCTTCATTGAGAACCCGGG-3' and 5'-CTCCGCTG TCACATAGTGGAAAACGTGAC-3'; for the *Cry1* E box, 5'-GCACGCGGGGCTGAG CCA-3' and 5'-CCGGTCCCGAGGCTGCCCG-3'; and for the first exon of *Cry1*, 5'-CTGCAACCAGCTCGGGCCGTC-3' and 5'-GATGAATGGGAGGCTGCCGAGGC-3'.

For semi-quantitative PCR analysis, PCR products were amplified for 20 cycles, resolved on 5% Tris-Borate-EDTA (TBE)-polyacrylamide gels, and Southern blotted. Blots were hybridized with ³²P-labelled oligonucleotides that were specific for the sequence of the amplified DNA fragments between the PCR primers. The PCR conditions were calibrated with known amounts of genomic DNA to confirm that the amplified products were in the linear range. Band intensity was quantified by phosphorimager analysis. The data were normalized to the input control, which consisted of PCR reactions from crosslinked chromatin before immunoprecipitation.

As negative controls, chromatin immunoprecipitations were performed in the absence of antibody or in the presence of an anti-haemagglutinin antibody (Santa Cruz Biotechnology). PCR products from these negative control samples were not detectable by ethidium bromide staining or Southern blot hybridization.

Immunoprecipitation assays

Immunoprecipitation was performed on nuclear extracts using anti-Clock antibody Clk-1-GP⁶. The immunoprecipitates were western blotted and probed with the following antibodies: anti-Clock antibody Clk-1-R⁶; anti-p300 monoclonal antibody (BD PharMingen; catalogue number 554215); anti-Cbp monoclonal antibody clone AC238 (NeoMarkers; catalogue number MS-696-PO); anti-Pcaf antibodies from Y. Nakatani and S. Berger; and anti-Gcn5 antibodies from Y. Nakatani and S. Berger.

Histone deacetylase assays

Immune complexes were assayed for HDAC activity as described¹⁹. The specificity of HDAC enzymes for specific lysine residues of individual histones observed *in vivo*²⁰ is not recapitulated faithfully *in vitro* using either free histones²¹ or chromatin²² as substrates.

Transcription start site determination

The major transcriptional start site of the *Cry1* gene was determined by primer extension, and the promoter region was sequenced. The +1 site is located at nucleotide 9947 of the mCG13749 Celera sequence.

Ligation-mediated PCR

We performed LM-PCR as described²³. An oligonucleotide specific to the 5' flanking region of *Cry1* (5'-GCGAGCC TCGGCCAATG-3') was annealed to 3.5 μ g of DNaseI-digested genomic DNA or *Cry1* plasmid DNA (control), and primer extension was performed with Sequenase (USB). A linker was ligated to the primer-extended DNA at 15 °C for 18 h, and the products were precipitated. The ligated products were subjected to a 20-cycle PCR reaction with Vent-DNA polymerase (New England Biolabs) using a nested primer (5'-GCCAATGGAGGCGGCTCCC-3') and a linker primer. The PCR products were templates for another round of Vent-PCR using a second nested, radiolabelled primer (5'-GCCAATGGAGGCGGCTCCCCG-3'). After 3 cycles, the samples were extracted, precipitated, and resolved on a sequencing gel. The nucleotide positions of the hypersensitive sites were determined by a sequencing ladder. Levels of products generated by LM-PCR were determined by phosphorimager analysis.

Electrophoretic mobility shift assays

The putative *Cry1* Rev-Erb/ROR probes were prepared by annealing the following complementary oligonucleotides: for *Cry1* site I, 5'-GTCCTTATAAAGTGGGTC-3' and 5'-CATAGCTGACCCAGTTTATAAGGAC-3'; for *Cry1*-mutant site I, 5'-GTCCTTAT AAAACCTTGGG-3' and 5'-CATAGCCCCAAGGTTTATAAGGAC-3'; for *Cry1* site II, 5'-AAAGAAACATGTAGGTCA-3' and 5'-TCGAAGATGACCTACATGTTTCTTT-3'; and for *Cry1* site III, 5'-TTATTTATATGGAGGTCA-3' and 5'-ATTGTTATGACCTCCATA TAAATAA-3'. Annealed oligonucleotides were labelled by filling the 5' overhang with α [³²P]dATP or α [³²P]dCTP with Klenow DNA polymerase. Proteins were expressed using TNT reticulocyte lysate kit (Promega). The protein-DNA interactions were carried out for 30 min at 4 °C with approximately 0.1 pmol of the double-stranded oligonucleotide in a reaction volume of 20 μ l containing 0.5% glycerol, 60 mM KCl, 1 μ g BSA, 30 mM HEPES (pH 7.5), 0.05% Triton X-100, and 60 ng of poly(dI-dC). Protein-DNA complexes were resolved on a 5% polyacrylamide gel electrophoresis (PAGE) for 3 h at 200 V, 4 °C, and visualized by autoradiography.

Luciferase reporter gene assays

Luciferase reporter gene assays were performed in COS-7 cells as described⁴. Tandem repeats of *Cry1* Rev-Erb/ROR binding sites were prepared by annealing the following oligonucleotides: for *Cry1* site I, 5'-CTAGCCTTATAAAGTGGGTGAGCTAGAAACT GGGTCAGCTATGAAACTGGGTGAGCTATA-3' and 5'-GATCTATAGCTGACCCAG TTTCATAGCTGACCCAGTTTCATAGCTGACCCAGTTTATAAGG-3'; for *Cry1*-mutant site I, 5'-CTAGCCTTATAAAGTGGGTGAGCTATGAAACTGGGTGAGCTATGAAACT CCTTGGGGCTATA-3' and 5'-GATCTATAGCCCCAAGGTTTTCATAGCCCCAAGG TTTCATAGCCCCAAGGTTTATAAGG-3'; for *Cry1* site II, 5'-ACATGTAGGTGATC TAAACATGTAGGTGATCTTATA-3' and 5'-GATCTATAAGATGACCTACATGTTTAGAT GACCTACATGTTTAGATGACCTACATGTTTAAAGG-3'; for *Cry1*-mutant site II, 5'-CTAGCCTTATAAAGTGGGTGCTAAAAACCTTTGGGTCTAAAAACCTTT GGGTCTTATA-3' and 5'-GATCTATAAGACCCAAAGGTTTTCATAGCCCCAAGG TTTTTCATAGCCCCAAGGTTTATAAGG-3'. The annealed oligonucleotides were phosphorylated and ligated into the pGL3-Promoter vector (Promega).

Received 10 September; accepted 25 November 2002; doi:10.1038/nature01314.

Published online 11 December 2002.

- King, D. P. *et al.* Positional cloning of the mouse circadian *Clock* gene. *Cell* **89**, 641–653 (1997).
- Bunger, M. K. *et al.* *Mop3* is an essential component of the master circadian pacemaker in mammals. *Cell* **103**, 1009–1017 (2000).
- Gekakis, N. *et al.* Role of the CLOCK protein in the mammalian circadian mechanism. *Science* **280**, 1564–1569 (1998).
- Jin, X. *et al.* A molecular mechanism regulating rhythmic output from the suprachiasmatic circadian clock. *Cell* **96**, 57–68 (1999).
- Kume, K. *et al.* mCRY1 and mCRY2 are essential components of the negative limb of the circadian clock feedback loop. *Cell* **98**, 193–205 (1999).
- Lee, C., Etcheberry, J. P., Cagampang, F. R., Loudon, A. S. & Reppert, S. M. Posttranslational mechanisms regulate the mammalian circadian clock. *Cell* **107**, 855–867 (2001).
- Reppert, S. M. & Weaver, D. R. Coordination of circadian timing in mammals. *Nature* **418**, 935–941 (2002).
- Akhtar, R. A. *et al.* Circadian cycling of the mouse liver transcriptome, as revealed by cDNA microarray, is driven by the suprachiasmatic nucleus. *Curr. Biol.* **12**, 540–550 (2001).
- Panda, S. *et al.* Coordinated transcription of key pathways in the mouse by the circadian clock. *Cell* **109**, 307–320 (2002).
- Storch, K. F. *et al.* Extensive and divergent circadian gene expression in liver and heart. *Nature* **417**, 78–83 (2002).
- Yagita, K., Tamanini, F., van Der Horst, G. T. & Okamura, H. Molecular mechanisms of the biological clock in cultured fibroblasts. *Science* **292**, 278–281 (2001).
- Orphanides, G. & Reinberg, D. A unified theory of gene expression. *Cell* **108**, 439–451 (2002).
- Jenuwein, T. & Allis, C. D. Translating the histone code. *Science* **293**, 1074–1080 (2001).
- Fry, C. J. & Peterson, C. L. Chromatin remodeling enzymes: who's on first? *Curr. Biol.* **11**, R185–R197 (2001).
- Takahata, S. *et al.* Transactivation mechanisms of mouse clock transcription factors, mClock and mArnt. *Genes Cells* **5**, 739–747 (2000).
- Preitner, N. *et al.* The orphan nuclear receptor REV-ERB α controls circadian transcription within the positive limb of the mammalian circadian oscillator. *Cell* **110**, 251–260 (2002).
- Ueda, H. R. *et al.* A transcription factor response element for gene expression during circadian night. *Nature* **418**, 534–539 (2002).
- Crosio, C., Cermakian, N., Allis, C. D. & Sassone-Corsi, P. Light induced chromatin modification in cells of the mammalian circadian clock. *Nature Neurosci.* **3**, 1241–1247 (2000).
- Wade, P. A., Jones, P. L., Vermaak, D. & Wolffe, A. P. Purification of a histone deacetylase complex

- from *Xenopus laevis*: preparation of substrates and assay procedures. *Methods Enzymol.* **304**, 715–725 (1999).
20. Rundlett, S. E. *et al.* HDA1 and RPD3 are members of distinct yeast histone deacetylase complexes that regulate silencing and transcription. *Proc. Natl Acad. Sci. USA* **93**, 14503–14508 (1996).
 21. Carmen, A. A., Rundlett, S. E. & Grunstein, M. HDA1 and HDA3 are components of a yeast histone deacetylase (HDA) complex. *J. Biol. Chem.* **271**, 15837–15844 (1996).
 22. Guschin, D., Wade, P. A., Kikyo, N. & Wolffe, A. P. ATP-dependent histone octamer mobilization and histone deacetylation mediated by the Mi-2 chromatin remodeling complex. *Biochemistry* **39**, 5238–5245 (2000).
 23. Pfeifer, G. P., Chen, H. H., Komura, J. & Riggs, A. D. Chromatin structure analysis by ligation-mediated and terminal transferase-mediated polymerase chain reaction. *Methods Enzymol.* **304**, 548 (1999).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements We thank D. Reinberg for the anti-RPB1 antibody; V. Sartorelli for the p300 expression construct; U. Schibler for Rev-Erbox reagents; H. R. Ueda for PATSER sequence analysis; and D. R. Weaver and C. L. Peterson for suggestions. This work was supported by grants from the NIH and the Defense Advanced Research Projects Agency (DARPA).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.M.R. (e-mail: steven.reppert@umassmed.edu).

IGF-1 receptor regulates lifespan and resistance to oxidative stress in mice

Martin Holzenberger*, Joëlle Dupont†, Bertrand Ducos*, Patricia Leneuve*, Alain Gélouin‡, Patrick C. Even§, Pascale Cervera|| & Yves Le Bouc*

* Institut National de la Santé et de la Recherche Médicale U515, and || Service d'Anatomie et de Cytologie Pathologiques, Hôpital Saint-Antoine, 75571 Paris 12, France

† Institut National de la Recherche Agronomique, 37380 Nouzilly, France

‡ Institut National de la Santé et de la Recherche Médicale U352, INSA, 69621 Villeurbanne, France

§ Institut National de la Recherche Agronomique, INA P-G, 75231 Paris 5, France

Studies in invertebrates have led to the identification of a number of genes that regulate lifespan, some of which encode components of the insulin or insulin-like signalling pathways^{1–3}. Examples include the related tyrosine kinase receptors InR (*Drosophila melanogaster*) and DAF-2 (*Caenorhabditis elegans*) that are homologues of the mammalian insulin-like growth factor type 1 receptor (IGF-1R). To investigate whether IGF-1R also controls longevity in mammals, we inactivated the IGF-1R gene in mice (*Igf1r*). Here, using heterozygous knockout mice because null mutants are not viable, we report that *Igf1r*^{+/-} mice live on average 26% longer than their wild-type littermates ($P < 0.02$). Female *Igf1r*^{+/-} mice live 33% longer than wild-type females ($P < 0.001$), whereas the equivalent male mice show an increase in lifespan of 16%, which is not statistically significant. Long-lived *Igf1r*^{+/-} mice do not develop dwarfism, their energy metabolism is normal, and their nutrient uptake, physical activity, fertility and reproduction are unaffected. The *Igf1r*^{+/-} mice display greater resistance to oxidative stress, a known determinant of ageing. These results indicate that the IGF-1 receptor may be a central regulator of mammalian lifespan.

Insulin/insulin-like signalling molecules that have been linked to longevity include DAF-2 and InR, and inactivation of the corresponding genes leads to increased lifespan in nematodes^{4,5} and

insects^{6,7}, respectively. Null mutations of the insect insulin-receptor substrate Chico, which acts downstream from InR, also extends lifespan⁸. Most long-lived *daf-2* and *Inr* mutants develop dwarfism and hypofertility; however, some *Inr* mutants display normal fertility and growth. This suggests that longevity may be regulated independently of body size and reproduction^{7,8}. DAF-2 and InR are structural homologues of a family of vertebrate tyrosine kinase receptors that includes the insulin receptor and the insulin-like growth factor type 1 receptor (IGF-1R). In vertebrates, the insulin receptor regulates energy metabolism⁹ whereas IGF-1R promotes growth¹⁰. IGF-1R is activated by its ligand IGF-I, which is secreted in response to growth hormone. It is unclear whether the insulin receptor or IGF-1R, or both, have taken over responsibility for lifespan regulation in vertebrates^{3,8}. The phenotypes of long-lived spontaneous mouse mutants studied so far indicate a probable link between longevity and growth. The long-lived *Prop1*^{df/df} (Ames dwarf) and *Pit1*^{dw/dw} (Snell dwarf) mutants^{11,12} display impaired pituitary gland development and low levels of pituitary hormones, including growth hormone. These mutants are sterile dwarfs. The recent targeted inactivation of the growth hormone receptor itself,

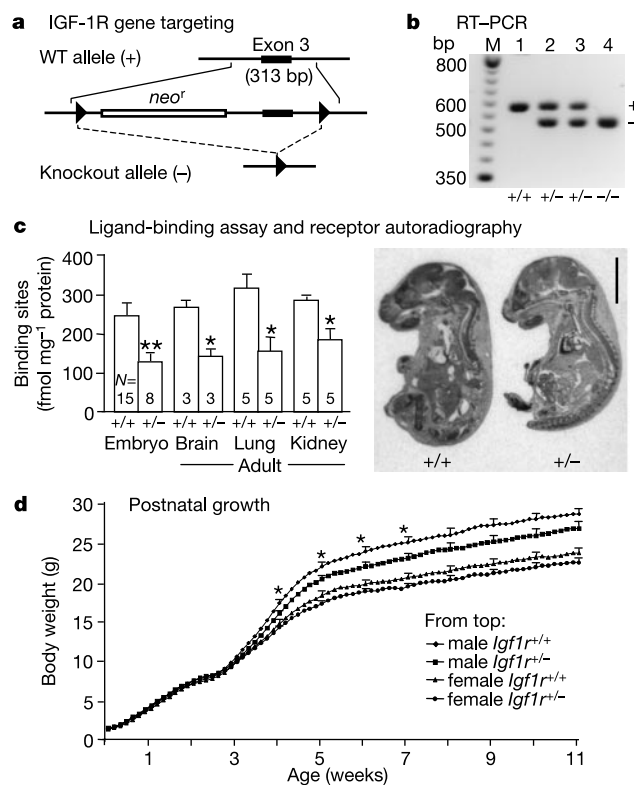


Figure 1 IGF-1R gene targeting, receptor expression and growth phenotype. **a**, We flanked exon 3 of the wild-type (WT) IGF-1R gene with a neomycin-resistance cassette (*neo*^r) and two *loxP* sites (triangles). Exon 3 and *neo*^r were then deleted by *Cre-lox* recombination, producing the *Igf1r*^{-/-} (knockout) allele^{14,15}. **b**, Allele-specific RT-PCR²⁵ revealed that heterozygous *Igf1r*^{+/-} mice produced mRNA from wild-type (+) and knockout (-) alleles (double band in lanes 2 and 3). M, DNA size marker. **c**, Although levels of IGF-1R were halved in *Igf1r*^{+/-} animals (bar graph), the relative distributions (autoradiographic pattern) were unchanged. Receptors were undetectable in *Igf1r*^{-/-} embryos (data not shown). Together, this indicates that the remaining, intact allele does not compensate for its inactivated homologue. Nonspecific binding was 8%. Scale bar, 5 mm. Asterisk, $P < 0.05$; double asterisk, $P < 0.01$ (Mann-Whitney *U*-test). See also Supplementary Information. **d**, *Igf1r*^{+/-} and wild-type siblings show identical growth until day 20. Thereafter, during the prepubescent growth spurt (weeks 3–5), slight deficits appear. Asterisk, $P < 0.05$ (Mann-Whitney *U*-test).

which strongly decreases circulating IGF-I and impairs growth and development, also increases lifespan¹³. Similarly, caloric restriction, the only efficient treatment known to increase mammalian lifespan, invariably reduces circulating IGF-I levels, and, if begun in juveniles, also engenders dwarfism. These findings led us to investigate whether mammalian lifespan is regulated by IGF-1R.

We inactivated the IGF-1R gene by homologous recombination using the *Cre-lox* strategy to delete the essential exon 3 of the gene^{14,15} (Fig. 1a). We found that homozygous null mutants (*Igf1r*^{-/-}) died at birth, as previously described in a study using classical insertional mutagenesis¹⁶. Our *Igf1r*^{+/-} mutant transmitted the null allele in the expected mendelian ratio (52%, *n* = 241). Wild-type and *Igf1r*-null transcripts were present (Fig. 1b); however, as *Igf1r*-null transcripts cannot be translated into functional protein¹⁵, IGF-1 receptor levels in *Igf1r*^{+/-} mice were half those in wild type *Igf1r*^{+/+} mice (Fig. 1c). Weight at birth and during the first three weeks of growth were nevertheless normal (Fig. 1d). Only after the natural weaning period (around day 20) did *Igf1r*^{+/-} males develop a modest, 8% growth deficit with respect to their *Igf1r*^{+/+} littermates (23.1 ± 0.7 g compared with 25.1 ± 0.7 g at age 7 weeks, $P < 0.05$), whereas the growth deficit did not exceed 6% in females (19.5 ± 0.6 g compared with 20.7 ± 0.5 g at 7 weeks, not significant (NS)). These modest weight differences affected all tissues to similar degrees, persisted throughout life (data not shown), and resembled the growth pattern observed in previous models of partial inactivation of the IGF-1R gene¹⁵. Thus, the bi-allelically expressed mouse IGF-1R gene is heteroinsufficient.

Fed *ad libitum* on a standard diet and maintained in regular housing until natural death, mice with only one functional IGF-1R allele significantly outlived their wild-type littermates (Fig. 2). *Igf1r*^{+/-} mice lived a mean of 26% longer than *Igf1r*^{+/+} controls ($P < 0.02$; Cox's test). If the sexes were evaluated separately, mutant females were found to live 33% longer than wild-type females ($P < 0.001$), whereas mutant males lived only 15.9% longer than

control males (NS). On average, *Igf1r*^{+/-} females outlived *Igf1r*^{+/-} males, whereas the opposite is normally the case in wild-type populations of the 129/J genetic background¹⁷ (Fig. 2). Thus, in *Igf1r*^{+/-} mice, the degree to which lifespan is extended depends on sex, as has been described for *Drosophila* mutants with impaired insulin-like signalling^{6,8}. Our ageing cohorts had a 5% tumour incidence, unrelated to genotype and consistent with the low (7%) general tumour incidence of mice with the 129/J background^{18,19}. Necropsy revealed a number of different diseases, consistent with the results of previous studies, showing that mice with this background do not develop specific age-related diseases¹⁹. We did not observe accidental deaths, although some of the sporadic mortality of younger males may have been consequences of fights for dominance.

Blood parameters (see Methods) were normal. However, serum IGF-I levels were upregulated in adult *Igf1r*^{+/-} mice (males, 795 ± 64 compared with 625 ± 30 ng ml⁻¹, $P < 0.01$; females, 716 ± 39 compared with 516 ± 14 ng ml⁻¹, $P < 0.001$; *n* = 8–10 per group) and may reflect an endocrine response to the reduced availability of IGF-1R¹⁵. Blood glucose levels in mice that were deprived of food overnight were unaffected. In fed animals, however, *Igf1r*^{+/-} males tended to have higher (+12%) and *Igf1r*^{+/-} females to have lower (-4.4%) blood glucose levels than the controls. Non-fasting insulin levels were nevertheless normal (males, 1.65 ± 0.12 compared with 1.40 ± 0.14 ng ml⁻¹; females, 1.58 ± 0.25 compared with 1.90 ± 0.25 ng ml⁻¹; *n* = 8–10 per group). We then tested glucose tolerance in mice that were deprived of food overnight (Fig. 3a, b) and found that *Igf1r*^{+/-} males had a significantly stronger glucose response than controls ($P < 0.001$). In *Igf1r*^{+/-} females, the response was slightly weaker than in wild-

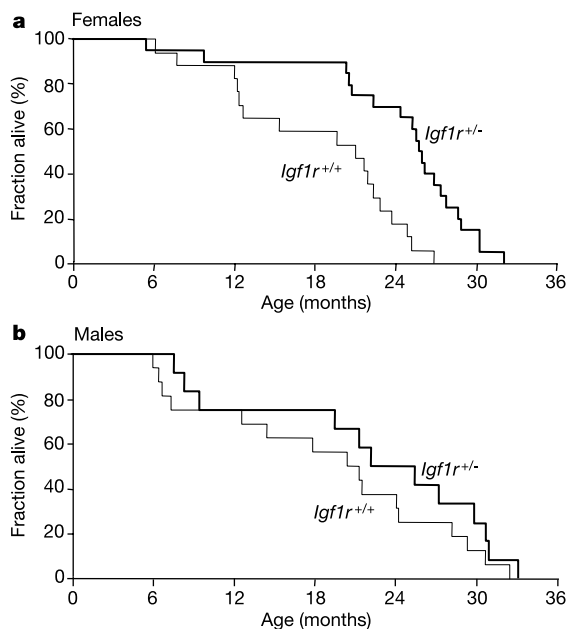


Figure 2 Lifespan extension in *Igf1r*^{+/-} mice with respect to *Igf1r*^{+/+} (WT) mice. **a**, *Igf1r*^{+/-} females (thick line) live a mean of 33% longer than their wild-type littermates (756 ± 46 compared with 568 ± 49 days; $P < 0.01$, *t*-test). Kaplan-Meier analysis of survival revealed a later decline in *Igf1r*^{+/-} mice compared with wild type ($P < 0.001$, Cox's test). **b**, *Igf1r*^{+/-} males live 15.9% longer than wild-type littermates (679 ± 80 compared with 585 ± 69 days; NS).

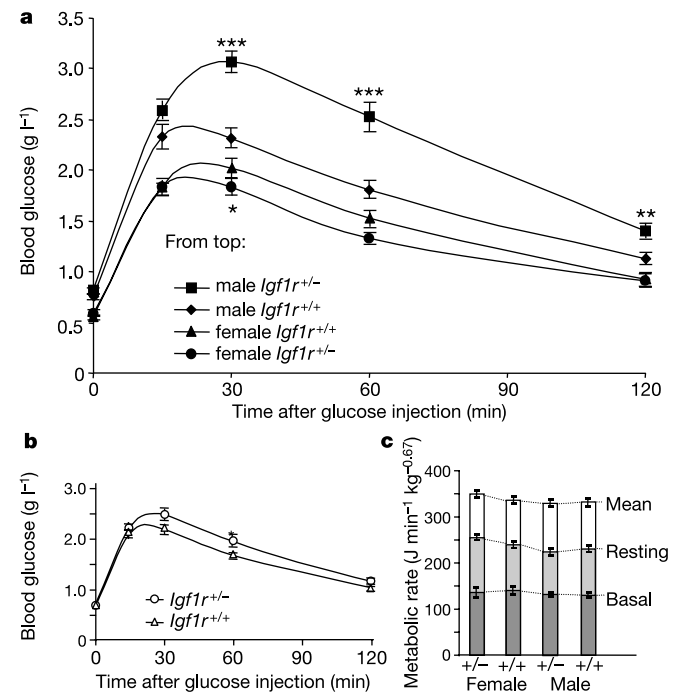


Figure 3 Glucose tolerance and energy metabolism in *Igf1r*^{+/-} mice. **a**, After an intraperitoneal glucose injection, the glucose response is strongest in mutant males. Note that the significant sex-related dimorphism of the response observed in wild-type mice is even more marked between mutants. Asterisk, $P < 0.05$; double asterisk, $P < 0.01$; triple asterisk, $P < 0.001$. **b**, Combining data from both sexes largely cancels out the male hyperglycaemic phenotype. Asterisk, $P < 0.05$; *n* = 89. **c**, Metabolic rate, measured by indirect calorimetry, did not differ between groups whether we determined mean (24 h), resting or basal metabolic rate (*n* = 11–12 per group).

type females ($P < 0.05$). It remains unclear whether this hyperglycaemic effect in *Igf1r*^{+/-} males is related to the reportedly compromised β -cell mass in these mice²⁰.

As metabolism may have an important role in ageing, we explored energy expenditure in *Igf1r*^{+/-} mice. Body temperature, indicative of metabolic activity and reported to be low in long-lived Ames dwarf mice²¹, was unaffected in *Igf1r*^{+/-} mice: skin surface temperature was $36.1 \pm 0.1^\circ\text{C}$ for all 4 groups ($n = 9$ –14 per group) and rectum temperature was 37.4 ± 0.2 compared with $37.5 \pm 0.2^\circ\text{C}$ in males and 37.7 ± 0.2 compared with $37.9 \pm 0.2^\circ\text{C}$ in females (NS). We analysed physical activity, using a photoelectric actimeter, and found identical circadian profiles for *Igf1r*^{+/-} and *Igf1r*^{+/+} mice (data not shown). Further-

more, as caloric restriction is known to extend rodent lifespan, we investigated the possibility that *Igf1r*^{+/-} mice were able to restrict their own food intake. We measured short-term (1–3 days) and long-term (90 days) food intake in adults. We observed only marginal differences between *Igf1r*^{+/-} and *Igf1r*^{+/+} mice, with mean food intake (in $\text{g d}^{-1}\text{kg}^{-1}$ body weight) being 190 ± 2 compared with 204 ± 4 (NS) in males and 148 ± 4 compared with 144 ± 2 (NS) in females, and mean water intake (in $\text{ml d}^{-1}\text{kg}^{-1}$ body weight) being 249 ± 7 compared with 255 ± 29 (NS) in males and 178 ± 4 compared with 166 ± 3 (NS) in females.

As mice may use nutrients with variable efficiency, we determined their metabolic rates. We obtained similar mean metabolic rates in fed animals over 24 h (males, 332.8 ± 6.7 compared with $336.0 \pm 8.4 \text{ J min}^{-1}\text{kg}^{-0.67}$; females, 351.9 ± 6.0 compared with $337.9 \pm 7.0 \text{ J min}^{-1}\text{kg}^{-0.67}$; NS, $n = 11$ –12 per group) (Fig. 3c). The resting metabolic rate was also similar between mutants and controls. As the difference between resting and 24-h metabolic rate depends mainly on physical activity, these results are consistent with the observed identical activity profiles. Even basal metabolic rate, measured in the fasted state, did not differ between mutants and controls (males, 133.2 ± 5.3 compared with $132.0 \pm 4.7 \text{ J min}^{-1}\text{kg}^{-0.67}$; females, 135.8 ± 9.4 compared with $140.9 \pm 6.6 \text{ J min}^{-1}\text{kg}^{-0.67}$).

Long-lived *C. elegans* *daf-2* mutants and long-lived dwarf mice display changes in fertility. We therefore monitored fertility and reproduction in *Igf1r*^{+/-} female mice from puberty to the age of 13 months. *Igf1r*^{+/-} females became fertile at 5.2 ± 0.1 weeks whereas *Igf1r*^{+/+} females became fertile at 5.6 ± 0.3 weeks (5.8 ± 0.1 compared with 6.2 ± 0.3 in males; $n = 7$ –11 per group). Although these differences were not significant, *Igf1r*^{+/-} mice became fertile,

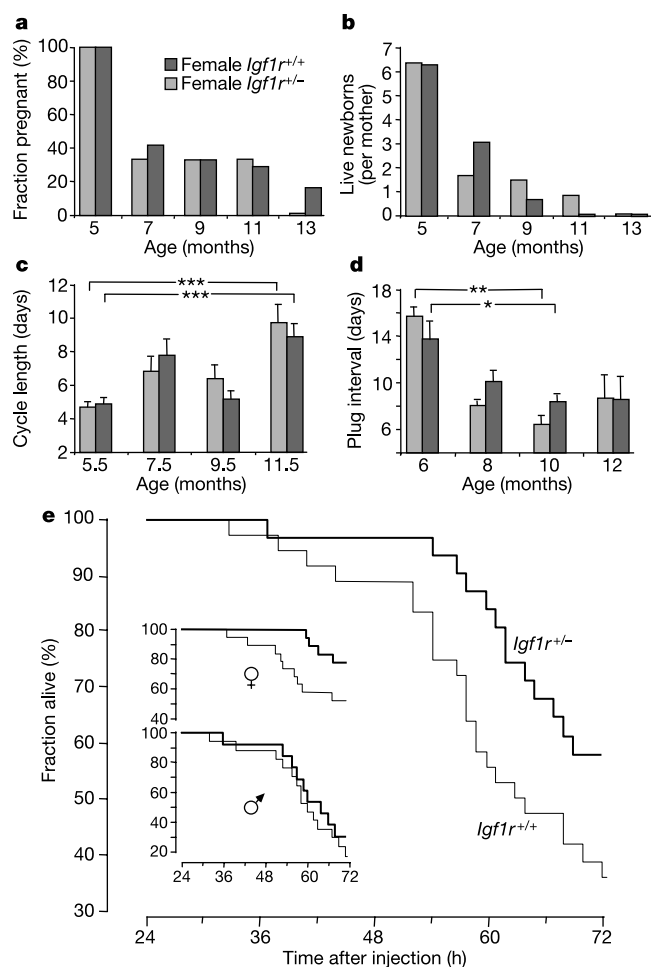


Figure 4 Mutants show normal fertility and are resistant to oxidative stress. **a**, Three-week mating results in a similar proportion of pregnancies in mutant and wild-type females. These proportions decline from 5 to 13 months of age ($P < 0.001$, χ^2 test). **b**, Although offspring decrease markedly with age, there are no consistent differences between *Igf1r*^{+/-} and control females. **c**, Oestrus cycle length increases significantly with age (triple asterisk, $P < 0.005$), reflecting changes in the hormonal control of ovarian function, but we observed no differences between genotypes. **d**, The interval between the first copulation plug (from a sterile male) and the next, indicative of ovarian capacity to maintain pseudogestation, decreases significantly with age in both groups. Asterisk, $P < 0.05$; double asterisk, $P < 0.02$; $n = 15$. **e**, Oxidative stress is induced by intraperitoneal paraquat injection (70 mg per kg body mass). We checked the mice every 2 h and censored the test at 72 h. Kaplan–Meier analysis shows significantly more survivors among *Igf1r*^{+/-} mice ($P < 0.05$, Cox's test; $n = 67$). When evaluated separately (inset), female mutants exhibit increased stress resistance ($P = 0.05$, log-rank test; $n = 37$), whereas the increase in males ($n = 30$) is small.

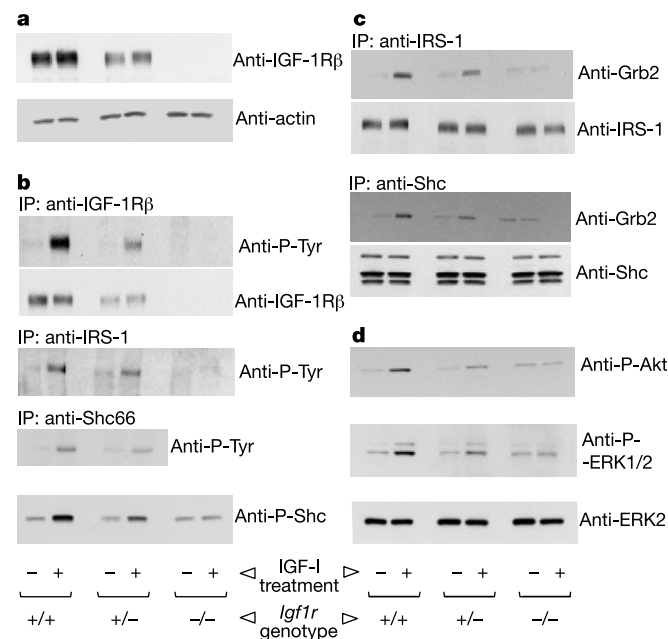


Figure 5 Lack of IGF-1R reduces activation of major intracellular signalling pathways in cultured MEFs. We stimulated MEFs with rhIGF-I (+) and analysed them by western blot. **a**, IGF-1R was reduced in *Igf1r*^{+/-} and absent from *Igf1r*^{-/-} MEFs. **b**, Anti-IGF-1R β , anti-IRS-1 and anti-p66 Shc immunoprecipitates (IP) were probed with anti-phosphotyrosine (anti-P-Tyr) antibodies. Phosphorylation of these proteins was reduced in *Igf1r*^{+/-} and absent from *Igf1r*^{-/-} MEFs. A phospho-Shc antibody revealed reduced activation of p52 Shc. **c**, **d**, IGF-I-induced association of Grb2 with IRS-1 or p52 Shc, and activation of ERK1/2 MAP kinases and Akt, are also reduced in *Igf1r*^{+/-} and absent from *Igf1r*^{-/-} MEFs.

on average, 3 days earlier than controls. This suggests that IGF-1R insufficiency does not delay sexual maturation, in contrast to the growth hormone receptor and binding protein (GHR/BP) knockout¹³. The litter size of young *Igflr*^{+/-} females was 6.3 ± 0.5 ($n = 7$ litters), which is not different from wild-type 129/Sv females (6.4 ± 0.2 live newborns; $n = 120$ litters). We analysed the age-related decline in female fertility, determining frequency of pregnancies, number of live newborns, mating behaviour, oestrus cycle length and ovarian capacity to maintain pseudogestation (Fig. 4a–d). As expected, fertility decreased markedly with age, but *Igflr*^{+/-} females and their controls displayed indistinguishable profiles.

Oxidative stress is a principal cause of ageing²², and mouse and fly mutants with enhanced resistance to oxidative stress are long-lived^{23,24}. We therefore subjected adult mice to oxidative stress by injection of paraquat, a herbicide that induces formation of reactive oxygen species. *Igflr*^{+/-} mutants resisted this challenge significantly longer than controls (Fig. 4e). This increase in stress resistance seemed to be more pronounced in female than in male mutants (Fig. 4e, inset). To further substantiate these results, we induced oxidative damage in cultured mouse embryonic fibroblasts (MEFs) by low concentrations of H₂O₂, and found that the proportion of surviving cells was significantly higher in *Igflr*^{+/-} than in control MEFs after 24- and 72-h treatments (24 h, $94\% \pm 3$ compared with $82\% \pm 3$, $P < 0.02$; 72 h, $88\% \pm 3$ compared with $68\% \pm 2$, $P < 0.001$).

Mutations of the *Drosophila* IRS homologue Chico⁸ and of other proteins acting downstream of IGF-1R, such as mouse p66 Shc²³ and *C. elegans* phosphatidylinositol-3-OH kinase (AGE-1)¹, but also regulation of forkhead transcription factor DAF-16 by Akt, increase lifespan. We therefore investigated how the reduced IGF-1R levels affected intracellular signalling in this model. We derived embryonic fibroblasts from wild type, *Igflr*^{+/-} and *Igflr*^{-/-} mice^{14,15} and studied signalling by western blot. As expected, *Igflr*^{+/-} cells showed 50% reduction in IGF-1R levels (Fig. 5a). Immunoprecipitation and western analysis also showed a marked reduction in IGF-I-induced tyrosine phosphorylation of IGF-1R and of its substrate IRS-1 (Fig. 5b). The tyrosine phosphorylation of both p52 and p66 isoforms of Shc, another major substrate of IGF-1R, were also reduced by half (Fig. 5b). This is of interest as reduced p66 Shc activation in *Igflr*^{+/-} cells could be a mechanism by which IGF-I regulates oxidative stress resistance. Shc and IRS-1 bind Grb2 on phosphorylation and thereby activate the mitogen-activated protein (MAP) kinase pathway, involved in mitogenic response. The amount of Grb2 co-immunoprecipitating with p52 Shc or IRS-1 are reduced to half in *Igflr*^{+/-} cells (Fig. 5c). Two important pathways activated by IGF-I are the MAP kinase ERK1/2 and phosphatidylinositol-3-OH kinase/Akt kinase signalling cascades. Consistently, IGF-I-stimulated phosphorylation of ERK1/2 and Akt is reduced by 40–50% in mutant cells (Fig. 5d). Together, this suggests that IGF-1R heteroinsufficiency downregulates the principal pathways stimulated by IGF-I.

These results show that a general decrease in IGF-1 receptor levels can increase lifespan in a mammalian species. Thus, the genetic link between insulin-like signalling and longevity, originally discovered in non-vertebrates^{4–6}, also seems to exist in higher vertebrates. Unlike long-lived mouse mutants with hypopituitarism^{11,12,25} or with complete lack of GHR/BP¹³, long-lived *Igflr*^{+/-} mutants, in which receptor levels were only reduced by 50%, did not develop dwarfism or hypofertility. We obtained these results using a 129/Sv genetic background. Preliminary results, however, using an IGF-1R knockdown mutation¹⁵ on a hybrid background (129/Sv × C57Bl/6J) confirm our findings on lifespan extension (M.H., unpublished data). IGF-1R is involved in the regulation of carbohydrate metabolism and in the pancreatic control of glucose homeostasis²⁶. It is therefore not surprising that we found abnormal regulation of blood glucose in *Igflr*^{+/-} mice. This abnormality affects only

males and is clearly a sex-related dimorphism. Reduced glucose tolerance is a symptom of prediabetes, and its potential consequences may have masked an otherwise possibly greater life-prolonging effect of IGF-1R insufficiency in males. We have observed previously other gender-related phenotypic differences in IGF-1R mutants^{14,15}, and have proposed the interplay of sex-dimorphic pulsatile growth hormone regulation, paracrine secretion and signalling of IGF-I, and androgen/oestrogen actions at the target cell level as possible explanations. Furthermore, similar sex dimorphism of longevity has been reported in *InR* mutant *Drosophila* and in long-lived Ames dwarf mice^{6,12}. It is tempting to speculate on the physiological significance of this sex dimorphism because major sex differences in lifespan have been found in numerous species, but clearly additional studies are needed. We thought that the observed variations in blood glucose might have had consequences for energy metabolism and expenditure, and that these mice might present features of caloric restriction. However, we show that *Igflr*^{+/-} mice have normal food uptake, physical activity or metabolic rate, which excludes metabolic differences as the cause of their longevity. It is, on the contrary, possible that the life-prolonging effects of caloric restriction are due to decreases in circulating IGF-I levels, mimicking the IGF-1R insufficiency produced here.

p66 Shc^{-/-} is the only other targeted mutation in mammals described so far that leads to a comparable increase in lifespan without inducing major side effects²³. The p66 isoform of Shc mediates cellular responses to oxidative stress and is, together with IRS-1, a major cytoplasmic signal transduction molecule for IGF-1R. Thus, the resistance of *Igflr*^{+/-} mice to oxidative stress is of considerable interest, and by showing that the stress-regulating p66 Shc is underphosphorylated in IGF-1R deficiency we found a plausible mechanism connecting IGF signalling to oxidative stress. Caloric restriction and decreases in the response to oxidative stress and in insulin-like growth factor signalling all efficiently extend lifespan in mice. However, it is unclear how these mechanisms cooperate and the extent to which they are independent. Data from *Drosophila chico*¹ mutants showing that lifespan is extended much more than can be explained by the modest increase in resistance to oxidative stress⁸ suggested that the two mechanisms operate independently, at least in part, to generate longevity. However, the issue of cooperation and independence of caloric restriction, insulin-like signalling and oxidative stress in lifespan extension remains unclear^{25,27}. Owing to strong oxidative stress-resistance phenotypes associated with *C. elegans* longevity mutations²⁸, these aspects merit further study in vertebrates. Our *Igflr*^{+/-} mutants provide an invaluable tool for future exploration of the mechanisms of lifespan regulation. However, it will also be necessary to try to overproduce IGF antagonists, to administer inhibitors of IGF-1R activation, or to block signal transduction. It has recently been shown that lifespan regulation through insulin-like signals in non-vertebrates probably occurs in a non-cell-autonomous fashion²⁹. Neurons in the central nervous system, by sensing the circulating levels of ligand, may have a central function in regulating the ageing of other tissues by means of hypothetical endocrine mechanisms. We have begun to investigate this possibility in a mammalian model, using the *Cre-lox* approach to produce brain-specific IGF-1R knockout mice. Homozygous mice for this mutation are microcephalic, sterile, and have a complex neuroendocrine dysfunction, but heterozygous mice are healthy and are useful for lifespan studies. □

Methods

Mice

The *Igflr* mutant, described elsewhere¹⁴ and available from <http://www.emma.rm.cnr.it>, was maintained in the heterozygous state in the 129/Sv genetic background. By mating *Igflr*^{+/-} males with 9- to 12-week-old 129/Sv wild-type females, we generated three cohorts, each composed of heterozygous *Igflr*^{+/-} and *Igflr*^{+/+} (wild type) littermates. Animals lived in conventional conditions: 23 °C, 14/10-h light/dark cycle, standard diet

(49% carbohydrates, 24% proteins, 5% lipids, 12% humidity, 10% minerals and fibre) and water *ad libitum*. We separated mice from mothers on day 30 and grouped them as 6 males or 6 females per cage, with both genotypes present in each cage. Mice from cohort 1 (20 *Igf1r*^{+/-} and 17 *Igf1r*^{+/+} females; 12 *Igf1r*^{+/-} and 16 *Igf1r*^{+/+} males) were checked daily but otherwise left undisturbed until they died naturally. Single surviving females were placed in the neighbouring cage, whereas single surviving males received a female for company. We performed necropsy whenever possible, including tumour immunohistochemistry. Four animals were killed when death appeared imminent, to reduce suffering. We drew Kaplan–Meier survival curves using dates of birth and death. Cohort 2 (9 *Igf1r*^{+/-} and 15 *Igf1r*^{+/+} females; 14 *Igf1r*^{+/-} and 11 *Igf1r*^{+/+} males) was used for blood biochemistry and analysis of glucose tolerance, food consumption, fertility and body composition. In cohort 3 (17 *Igf1r*^{+/-} and 11 *Igf1r*^{+/+} females, 14 *Igf1r*^{+/-} and 15 *Igf1r*^{+/+} males) we analysed growth, energy expenditure (by indirect calorimetry), blood parameters, glucose tolerance, and finally *in vivo* resistance to oxidative stress induced by methyl viologen (paraquat) injection. We conducted experiments according to institutional guidelines for care of laboratory animals.

Igf-1 receptor expression

Recombinant human IGF-I (rhIGF-I, for *in vitro* ligand-binding assays, described elsewhere^{14,15}) and rhdes(1-3)IGF-I (for autoradiography) were labelled with ¹²⁵I (see also Supplementary Information).

For the allele-specific expression assay, we used total RNA from *Igf1r*^{+/-} embryos and triplex reverse transcriptase-polymerase chain reaction (RT-PCR)²⁶. We co-amplified corresponding fragments from the coding region of the messenger RNA specific for the wild-type and the inactivated receptor allele. The single forward primer 5'-CGCCTGAAAACTGCACG-3' annealed with exon 2, the first reverse primer 5'-AGCTGCCAGGCACTCCG-3' annealed with exon 3, and the second reverse primer 5'-GCAGGGGATACAGTACATGTTT-3' spanned the knockout-specific splice junction between exons 2 and 4. RT-PCR products of 518 base pairs (bp) corresponded to the transcript of the knockout allele, and 574-bp products corresponded to wild type. We extracted total RNA from embryo homogenates by RNAXEL and performed One-Step RT-PCR using GeneAmp 2400 cyclers. For the reverse transcription reaction, we incubated 30 ng total RNA at 50 °C for 30 min and at 94 °C for 2 min, followed by 40 PCR cycles, each consisting of 30-s segments at 94, 59 and 72 °C.

Postnatal growth

To synchronize individual growth, we recomposed 7 litters (cohort 3) on day 1 to yield 8 or 9 pups per mother. We identified newborn mice with coloured ink and permanently numbered them on day 8. For 11 weeks we weighed them daily at 16:00 on an electronic balance. Growth curves used sliding means of present weight and weight on the day before and after.

Fertility

We determined the onset of male and female fertility by mating *Igf1r*^{+/-} and *Igf1r*^{+/+} mice from day 30 onwards with fertile wild-type partners (three females per male). The age of delivery and litter size (when testing females) were compared between genotypes. To evaluate the decline in female fertility over time we used three sets of parameters. Measurements started at 5 months of age and were repeated 3 or 4 times at 2-month intervals. First, we monitored the oestrus cycle by vaginal smear histology for 18 days. Second, we analysed sexual behaviour and the duration of pseudogestation by mating females for 2 weeks with vasectomized males and recording vaginal plugs. Third, we mated females with fertile males for three weeks and recorded the resulting pregnancies and offspring.

Blood tests

We measured total bilirubin, cholesterol, creatinine, glucose, lactate, total protein, triglycerides, urea and uric acid in 5-month-old mice. Circulating IGF-I was measured using a double-antibody RIA from Diagnostic Systems Laboratories and plasma insulin using the Linco Sensitive Rat Insulin RIA. We tested glucose tolerance in mice that had gone without food for 14 h overnight by intraperitoneal injection with 2 g kg⁻¹ body weight of 25% D-glucose. We measured circulating glucose in tail blood at 0, 15, 30, 60 and 120 min using Lifescan Glucotouch.

Indirect calorimetry

We determined metabolic rate by indirect 24-h calorimetry (see Supplementary Information for details).

Experiments using MEFs

We established female MEFs from *Igf1r*^{+/+}, *Igf1r*^{+/-} and *Igf1r*^{-/-} embryonic day (E)14 fetuses. IGF signalling pathways and *in vitro* resistance to H₂O₂ were studied in early passages of MEFs. For IGF signalling we removed serum (10% FCS) from MEF cultures 16 h before analysis. Cells were stimulated with 3 nM rhIGF-I (Genentech) for 10 min before analysis, except for detection of phosphorylated Shc (5 min). For H₂O₂ resistance we treated MEFs with 100 μM H₂O₂ and determined cell viability after 1–3 days using Trypan blue and a haemocytometer. *n* = 6 for each group, in two independent experiments.

Western blotting

We performed immunoprecipitation and western blotting as described³⁰ (see Supplementary Information for details).

Statistics

For group comparisons, we used Student's *t*-test. Means are expressed ± standard error of the mean (s.e.m.). Error bars represent the s.e.m. We determined the significance of survival curves by Cox's test. We used nonparametric Mann–Whitney and χ² tests where indicated.

Received 4 October; accepted 18 November 2002; doi:10.1038/nature01298.

Published online 4 December 2002.

- Guarente, L. & Kenyon, C. Genetic pathways that regulate ageing in model organisms. *Nature* **408**, 255–262 (2000).
- Kenyon, C. A conserved regulatory system for aging. *Cell* **105**, 165–168 (2001).
- Gems, D. & Partridge, L. Insulin/IGF signalling and ageing: seeing the bigger picture. *Curr. Opin. Genet. Dev.* **11**, 287–292 (2001).
- Kimura, K. D., Tissenbaum, H. A., Liu, Y. & Ruvkun, G. Daf-2, an insulin receptor-like gene that regulates longevity and diapause in *Caenorhabditis elegans*. *Science* **277**, 942–946 (1997).
- Tissenbaum, H. A. & Ruvkun, G. An insulin-like signaling pathway affects both longevity and reproduction in *Caenorhabditis elegans*. *Genetics* **148**, 703–717 (1998).
- Tatar, M. *et al.* A mutant *Drosophila* insulin receptor homolog that extends life-span and impairs neuroendocrine function. *Science* **292**, 107–110 (2001).
- Bartke, A. Mutations prolong life in flies; implications for aging in mammals. *Trends Endocrinol. Metab.* **12**, 233–234 (2001).
- Clancy, D. J. *et al.* Extension of life-span by loss of CHICO, a *Drosophila* insulin receptor substrate protein. *Science* **292**, 104–106 (2001).
- Kulkarni, R. N. *et al.* Tissue-specific knockout of the insulin receptor in pancreatic beta cells creates an insulin secretory defect similar to that in type 2 diabetes. *Cell* **96**, 329–339 (1999).
- Lupu, L., Terwilliger, J. D., Lee, K., Segre, G. V. & Efstratiadis, A. Roles of growth hormone and insulin-like growth factor 1 in mouse postnatal growth. *Dev. Biol.* **229**, 141–162 (2001).
- Flurkey, K., Papaconstantinou, J., Miller, R. A. & Harrison, D. E. Lifespan extension and delayed immune and collagen aging in mutant mice with defects in growth hormone production. *Proc. Natl Acad. Sci. USA* **98**, 6736–6741 (2001).
- Brown-Borg, H. M., Borg, K. E., Meliska, C. J. & Bartke, A. Dwarf mice and the ageing process. *Nature* **384**, 33 (1996).
- Coschigano, K. T., Clemmons, D., Bellushi, L. L. & Kopchick, J. J. Assessment of growth parameters and life span of GHR/BP gene-disrupted mice. *Endocrinology* **141**, 2608–2613 (2000).
- Holzenberger, M. *et al.* Experimental IGF-I receptor deficiency generates a sexually dimorphic pattern of organ-specific growth deficits in mice, affecting fat tissue in particular. *Endocrinology* **142**, 4469–4478 (2001).
- Holzenberger, M. *et al.* A targeted partial invalidation of the insulin-like growth factor I receptor gene in mice causes a postnatal growth deficit. *Endocrinology* **141**, 2557–2566 (2000).
- Liu, J. P., Baker, J., Perkins, A. S., Robertson, E. J. & Efstratiadis, A. Mice carrying null mutations of the genes encoding insulin-like growth factor I (Igf-1) and type I IGF receptor (Igf1r). *Cell* **75**, 59–72 (1993).
- Storer, J. B. Longevity and gross pathology at death in 22 inbred mouse strains. *J. Gerontol.* **21**, 404–409 (1966).
- Stevens, L. C. & Little, C. C. Spontaneous testicular teratomas in an inbred strain of mice. *Proc. Natl Acad. Sci. USA* **40**, 1080–1087 (1954).
- Smith, G. S., Walford, R. L. & Mickey, M. R. Lifespan and incidence of cancer and other diseases in selected long-lived inbred mice and their F1 hybrids. *J. Natl Cancer Inst.* **50**, 1195–1213 (1973).
- Whithers, D. J. *et al.* Irs-2 coordinates IGF-1 receptor-mediated β-cell development and peripheral insulin signalling. *Nature Genet.* **23**, 32–39 (1999).
- Hunter, W. S., Croson, W. B., Bartke, A., Gentry, M. V. & Meliska, C. J. Low body temperature in long-lived Ames dwarf mice at rest and during stress. *Physiol. Behav.* **67**, 433–437 (1999).
- Finkel, T. & Holbrook, N. J. Oxidants, oxidative stress and the biology of ageing. *Nature* **408**, 239–247 (2000).
- Migliaccio, E. *et al.* The p66^{shc} adaptor protein controls oxidative stress response and life span in mammals. *Nature* **402**, 309–313 (1999).
- Sun, J. & Tower, J. FLP recombinase-mediated induction of Cu/Zn-superoxide dismutase transgene expression can extend the life span of adult *Drosophila melanogaster* flies. *Mol. Cell Biol.* **19**, 216–228 (1999).
- Bartke, A. *et al.* Extending the lifespan of long-lived mice. *Nature* **414**, 412 (2001).
- Kulkarni, R. N. *et al.* β-cell specific deletion of the IGF-1 receptor leads to hyperinsulinemia and glucose intolerance but does not alter β-cell mass. *Nature Genet.* **31**, 111–115 (2002).
- Clancy, D. J., Gems, D., Hafen, E., Leivers, S. J. & Partridge, L. Dietary restriction in long-lived dwarf flies. *Science* **296**, 319 (2002).
- Honda, Y. & Honda, S. Oxidative stress and life span determination in the nematode *Caenorhabditis elegans*. *Ann. NY Acad. Sci.* **959**, 466–474 (2002).
- Wolkow, C. A., Kimura, K. D., Lee, M. S. & Ruvkun, G. Regulation of *C. elegans* life-span by insulinlike signaling in the nervous system. *Science* **290**, 147–150 (2000).
- Dupont, J., Karas, M. & LeRoith, D. The potentiation of estrogen on insulin-like growth factor I action in MCF-7 human breast cancer cells includes cell cycle components. *J. Biol. Chem.* **275**, 35893–35901 (2000).

Supplementary Information accompanies the paper on Nature's website
(<http://www.nature.com/nature>).

Acknowledgements We thank P. Monget for contributions to the experimental design; N. R. Holzenberger for assistance with metabolic and growth studies; G. Hamard for help with MEFs; J. Sappa for language revision; F. Veinberg for blood biochemistry; and P. Casanovas and M.-C. Samson for animal care. MENRT sponsored this study with a grant to M.H. and Y.L.B. We thank D. LeRoith for support to J.D.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.H. (e-mail: holzenberger@st-antoine.inserm.fr).

Srb10/Cdk8 regulates yeast filamentous growth by phosphorylating the transcription factor Ste12

Chris Nelson, Susan Goto, Karen Lund, Wesley Hung & Ivan Sadowski

Department of Biochemistry and Molecular Biology, University of British Columbia, Vancouver, British Columbia V6T 1Z3, Canada

The budding yeast *Saccharomyces cerevisiae* differentiates into filamentous invasively growing forms under conditions of nutrient limitation^{1,2}. This response is dependent on the transcription factor Ste12 and on the mating pheromone-response mitogen-activated protein (MAP) kinase cascade¹, but a mechanism for regulation of Ste12 by nutrient limitation has not been defined. Here we show that Ste12 function in filamentous growth is regulated by the cyclin-dependent kinase Srb10 (also known as Cdk8), which is associated with the RNA polymerase II holoenzyme. Srb10 inhibits filamentous growth in cells growing in rich medium by phosphorylating Ste12 and decreasing its stability. Under conditions of limiting nitrogen, loss of Srb10 protein and kinase activity occurs, with a corresponding loss of Ste12 phosphorylation. Mutation of the Srb10-dependent phosphorylation sites increases pseudohyphal development but has no effect on the pheromone response of haploid yeast. Srb10 kinase activity is also regulated independently of the mating pheromone-response pathway. This indicates that Srb10 controls Ste12 activity for filamentous growth in response to nitrogen limitation and is consistent with the hypothesis that Srb10 regulates gene-specific activators in response to physiological signals to coordinate gene expression with growth potential.

Yeast differentiate into elongated invasively growing filamentous forms to facilitate foraging under conditions of nitrogen or carbon limitation^{1,2}. This response requires the transcription factors Ste12 and Tec1, which bind cooperatively to filamentous response elements (FREs) within the promoters of filamentous response genes³. Transcription from FREs is induced in response to nitrogen starvation⁴, and filamentous response requires upstream signalling components of the MAP kinase pheromone-response pathway, but a mechanism for regulation of Ste12 by nutrient limitation has not been identified.

Yeast lacking the RNA polymerase II holoenzyme-associated cyclin-dependent kinase (CDK) Srb10 show increased expression of Ste12-dependent filamentous response genes (ref. 5 and Fig. 1a), form more extensive pseudohyphae in nitrogen-depleted (SLAD) medium (Fig. 1b) and show constitutive pseudohyphae even on rich medium (data not shown). This effect is dependent on STE12, because diploid yeast with disruptions of both *srb10* and *ste12* do not form pseudohyphae on nitrogen-limiting medium (Fig. 1b, *srb10, ste12*) and have barely detectable levels of FRE-

dependent transcription (Fig. 1a). Disruption of *ste11*, which encodes the pheromone-responsive MAP/extracellular-signal-regulated kinase (ERK) kinase kinase (MEKK), also prevents filamentous growth (Fig. 1b) and FRE-dependent transcription (Fig. 1a) in *srb10* yeast. Therefore, elimination of Ste12 activity, by either gene disruption or impairing basal signalling, prevents the hyperfilamentous response of yeast lacking Srb10, suggesting that Srb10 may inhibit filamentous growth by modulating the activity of Ste12.

Consistent with this possibility, we found that Ste12 is a substrate for purified recombinant Srb10–Srb11 (Cdk8–cyclinC) complexes *in vitro* (Fig. 1c). Wild-type Srb10–Srb11 complexes phosphorylated recombinant Ste12 (Fig. 1c, lane 1), but not the inhibitors Dig1 and Dig2 (lanes 3 and 4). By contrast, a mutant Srb10 protein with impaired kinase activity (Asp290Ala)⁵ only weakly phosphorylated Ste12 *in vitro* and also underwent inefficient autophosphorylation (Fig. 1c, lane 2). Ste12 was phosphorylated on two peptides by Srb10 *in vitro* (Fig. 1d, peptides 3 and 4); these peptides co-migrated with phosphopeptides 3 and 4 derived from Ste12 labelled *in vivo* (ref. 6,

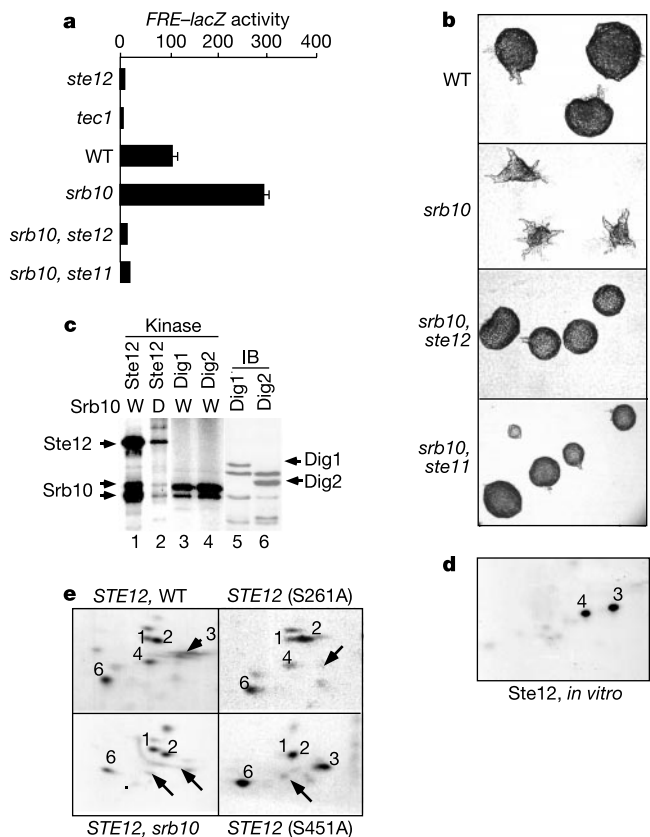


Figure 1 Phosphorylation of Ste12 by Srb10 inhibits filamentous responsive transcription. **a**, Deletion of *srb10* increases FRE-dependent transcription. Homozygous diploid strains bearing an *FRE-lacZ* reporter³ were grown at 30 °C in selective media and assayed for β -galactosidase activity⁹. **b**, Hyperfilamentation of *srb10* strains requires STE12 and an intact MAPK cascade. Homozygous diploid yeast were streaked on SLAD plates and photographed after 3 d. **c**, Srb10 phosphorylates Ste12 *in vitro*. Kinase reactions were carried out with purified recombinant wild-type Srb10–Srb11 (ref. 12; lanes 1, 3, 4) or a kinase-deficient Asp290Ala Srb10–Srb11 mutant (lane 2) plus recombinant Ste12 (lanes 1, 2), GST–Dig1 (lane 3) or GST–Dig2 (lane 4). Input substrate GST–Dig1 and GST–Dig2 proteins were immunoblotted with antibodies to GST (lanes 5, 6). **d**, Srb10 phosphorylates two sites on Ste12. Tryptic phosphopeptide analysis of *in vitro* phosphorylated Ste12 was done as described⁶. The resulting peptides co-migrate with Ste12 phosphopeptides 3 and 4 (not shown). **e**, Srb10 is required for phosphorylation of Ser 261 and Ser 451 in Ste12 *in vivo*. Shown is tryptic phosphopeptide analysis of *in vivo* phosphorylated Ste12 recovered from wild-type or *srb10* strains expressing wild-type STE12 (left panels), S261A STE12 (top right) or S451A STE12 (bottom right).

Fig. 1e, top left, and data not shown). Phosphorylation of these sites *in vivo* is dependent on *Srb10*, because phosphopeptides 3 and 4 were not derived from *Ste12* labelled in cells lacking *srb10* (Fig. 1e, compare left panels).

On the basis of their mobilities⁷, and by analysing deletion and point mutants, we determined that these phosphopeptides are produced by phosphorylation of *Ste12* at Ser 261 and Ser 451. Accordingly, phosphopeptide 3 was absent from an *Ser261Ala* mutant *Ste12* protein (*S261A Ste12*) labelled *in vivo* (Fig. 1e, top right). Although the peptide containing Ser 451 contains a second site of phosphorylation that can also produce phosphopeptide 4 (data not shown, but see below), phosphopeptide 4 was noticeably weaker on *Ste12* with a *Ser451Ala* (*S451A Ste12*) mutation (Fig. 1e, bottom right). Consistent with these results, recombinant *S261A Ste12* was not phosphorylated by *Srb10*–*Srb11* *in vitro*, whereas phosphorylation of recombinant *S451A Ste12* produced only phosphopeptide 3 (data not shown). These observations show that *Srb10* phosphorylates two sites in the activation domain^{8,9} of *Ste12* at Ser 261 and Ser 451.

Mutation of the *Srb10*-dependent phosphorylation sites on *Ste12* caused enhanced filamentous response in diploid yeast, as cells expressing either the *S261A Ste12* or the *S451A Ste12* mutants developed enhanced pseudohyphae as compared with wild type on nitrogen-limiting agar (Fig. 2a). In addition, *FRE*-dependent transcription was increased considerably in yeast expressing these mutants (Fig. 2b). In contrast to their negative effect on filamentous

growth, the phosphorylations at Ser 261 and Ser 451 do not regulate pheromone response. Mutations of these sites, either alone or in combination, did not affect induction of pheromone-responsive genes, as measured by a *FUS1-lacZ* reporter (Fig. 2c), nor did they affect sensitivity to pheromone (Fig. 2d). In addition, haploid yeast with disruptions in *srb10* did not have a mating defect (data not shown) and induced expression of *FUS1-lacZ* comparably to wild-type strains (Fig. 2c). Therefore, *Srb10*-dependent phosphorylation of *Ste12* specifically regulates *Ste12* function in pseudohyphal growth, but not pheromone response.

The inhibitory effect of *Srb10* on *Ste12* seems to be mediated primarily through regulation of protein stability. Pulse-chase labelling indicated that the half-life of wild-type *Ste12* protein is about 30 min in wild-type cells (Fig. 3a, b). Both disruption of *srb10* and the *S261A* and *S451A Ste12* mutations stabilized *Ste12* to beyond 90 min (Fig. 3a, b). We did not observe an effect of the *S261A* and *S451A* mutations on interaction of *Ste12* with glutathione *S*-transferase (GST) fusions of the known regulators for filamentous growth *Dig1*, *Dig2* or *Kss1*. Chromatin immunoprecipitation experiments also showed that phosphorylation does not affect interaction of *Ste12* with *FRE*-containing promoters *in vivo* (data not shown). These observations indicate that *Srb10* regulates *Ste12* function by controlling its stability.

This conclusion prompted us to consider how *Srb10*, and its phosphorylation of *Ste12*, can specifically inhibit pseudohyphal growth but not pheromone response. It has been shown that *STE12* messenger RNA expression is significantly lower in diploid cells

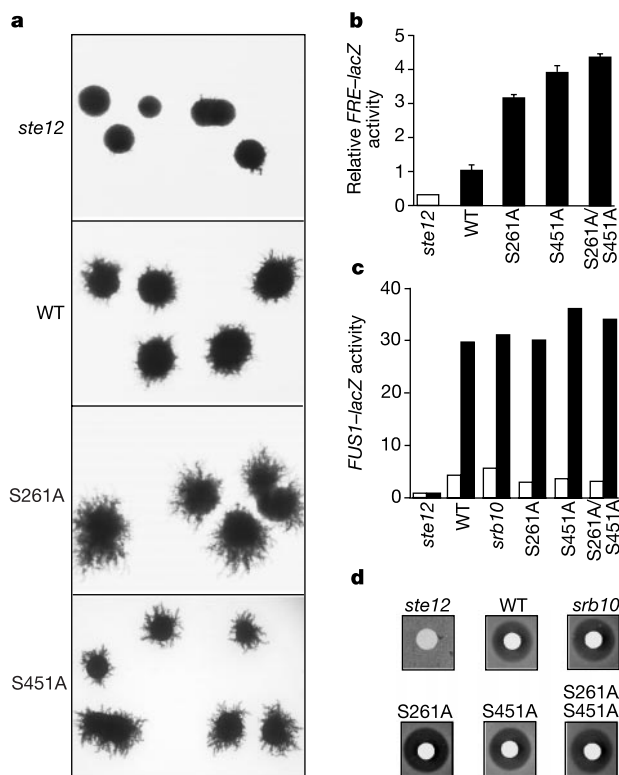


Figure 2 Phosphorylation of Ser 261 and Ser 451 of *Ste12* inhibits filamentous growth but not pheromone response. **a**, Homozygous diploid yeast expressing the indicated *STE12* alleles were grown on SLAD agar, and photos were taken after 3 d. **b**, Diploid yeast expressing the indicated *STE12* alleles were transformed with an *FRE(TEC1)-lacZ* reporter plasmid³, and β -galactosidase activity was measured after 4 h growth in SLAD medium. **c**, Either W303-1A MAT *a ste12* yeast carrying a vector control (*ste12*) or expressing the indicated *STE12* alleles (wild-type, *S261A*, *S451A* and *S261A/S451A*), or MAT *a ste12 srb10* (*srb10*) yeast were assayed for expression of a *FUS1-lacZ* reporter gene before (open bars) and 20 min after the addition of pheromone (filled bars). **d**, Pheromone-response halo assays were done on the strains described in **c** by spotting 20 μ g of synthetic α -factor onto sterile filter discs on a lawn of cells.

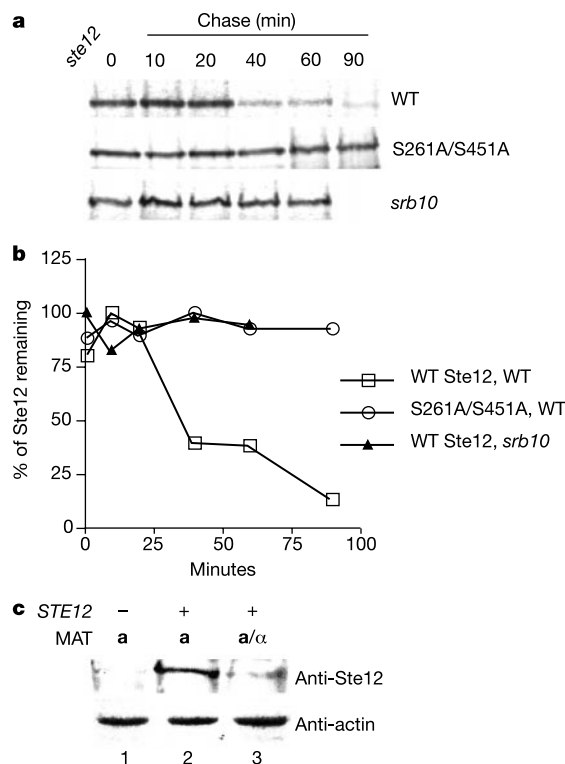


Figure 3 Stability of *Ste12* is controlled by *Srb10*. **a**, Pulse-chase analysis of *Ste12*. *ste12* or *ste12 srb10* (*srb10*) yeast carrying a vector control (*ste12*) or *GAL*-expression plasmids producing wild-type *STE12* or *S261A/S451A STE12* were labelled with [³⁵S]methionine in the presence of 2% galactose for 1 h. Samples were chased into YEPD medium plus 25 mM methionine, and *Ste12* was recovered by immunoprecipitation⁶ at the indicated times. **b**, Quantification of pulse-chase data. **c**, *Ste12* expression is reduced in diploid cells. Extracts were prepared from MAT *a ste12* (lane 1), wild-type MAT *a* (lane 2) and wild-type MAT *a/α* (lane 3) strains. Total protein (100 μ g) was resolved by SDS-PAGE, and proteins were detected by immunoblotting with antibodies to *Ste12* (top) or actin (bottom).

relative to haploids¹⁰. This difference is also reflected in the steady-state amounts of protein, because we observed 5–10 times less Ste12 in diploids relative to haploids (Fig. 3c, compare lanes 2 and 3), indicating that Ste12-regulated genes in diploids must be particularly sensitive to Ste12 stability. Consistent with this notion, double S261A and S451A Ste12 mutations caused only a subtle increase in haploid filamentous growth (data not shown).

Srb10 activity is inhibited when yeast are starved for nitrogen. Srb10 recovered by immunoprecipitation from cells growing in rich medium efficiently phosphorylated recombinant Ste12 protein *in vitro* (Fig. 4a, lane 2), whereas kinase activity was absent in immunoprecipitates from cells shifted to nitrogen-limiting SLAD medium for 4 h (Fig. 4a, lane 3). Nitrogen limitation decreased Srb10 abundance, as immunoblotting indicated that Srb10 was absent from cells after 4 h in SLAD medium (Fig. 4a, lane 3, HA-Srb10). Inhibition of Srb10 activity in nitrogen-starved cells was also observed on Ste12 phosphorylation *in vivo*. Ste12 recovered from cells labelled with ³²P in nitrogen-limiting conditions did not contain phosphopeptide 3 (Fig. 4b, centre). In addition, phosphopeptide 4, containing Ser 451, was shifted slightly in nitrogen-starved cells to produce a doublet spot when co-migrated with phosphopeptides from Ste12 labelled in rich medium (Fig. 4b, right, 4 and 4*). This effect was caused by an additional phosphorylation on the tryptic peptide containing Ser 451 that occurs independently of Srb10 (data not shown, but see Fig. 1e). These observations indicate that nitrogen limitation causes a loss of Srb10-dependent phosphorylation of Ste12 by a mechanism involving depletion of Srb10 protein.

From the data presented here, we propose a new model of the regulation of filamentous growth in yeast. We suggest that the main requirement of the upstream MAP kinase pathway in filamentous growth is to provide Ste12 with basal activity, allowing its interaction with components of the RNA polymerase II holoenzyme and phosphorylation by Srb10. Under ideal growth conditions, in which Srb10 is fully active, Ste12 activity is limited by its rapid turnover rate. Nitrogen limitation downregulates quantities of Srb10, which allows the accumulation of unphosphorylated, more stable Ste12 protein. In diploid cells, this can account for induction of FRE-

dependent genes where Ste12 protein may be limiting. Srb10 activity *in vivo* is not affected by disruptions of *ste11* or disruptions of both *kss1* and *fus3* (ref. 6 and data not shown). Consequently, activity of Srb10 must be regulated in response to nitrogen limitation by a signalling mechanism that is independent of the pheromone-response MAP kinase cascade. Regulation of several transcriptional activators by CDKs associated with the RNA polymerase II holoenzyme represents an efficient mechanism to disseminate signals generated by physiological changes. It is likely that additional factors involved in filamentous response, such as Tec1, Phd1, Flo8 and Ash1, might also be regulated by Srb10-dependent phosphorylations. Accordingly, we found that *srb10* disruptions cause a greater increase in haploid invasive growth than do mutations of the Srb10-dependent Ste12 phosphorylation sites (data not shown).

Our data support the hypothesis that Srb10 functions to regulate transcriptional activators in response to general nutritional and physiological conditions. Srb10 protein disappears on a switch to limiting nitrogen (Fig. 4a), and as cells pass the diauxic shift where nutrients become limiting⁵. In addition, *SRB10* was genetically identified as a regulator of the transition to stationary phase¹¹. Srb10 is known to also phosphorylate the yeast transactivators Gal4, Msn2 and Gcn4. Loss of Srb10 activity stabilizes both Ste12 and Gcn4, and allows nuclear accumulation of Msn2 (ref. 12), to activate transcription of stress-induced genes. By contrast, Srb10 is required for full activity of Gal4, and thus inhibition of Srb10 under nutrient-limiting conditions may abrogate expression of galactose-responsive genes in tune with the cell growth potential¹³. Considering the pivotal role of Srb10 in coordinating yeast morphological development, and its conservation among eukaryotes, it will be important to identify targets of CDK8 in mammalian cells. □

Methods

Yeast strains and plasmids

Yeast strains were derived from the Σ 1278b background unless indicated otherwise¹⁴. *STE12* alleles were introduced using derivatives of the plasmid pLS172, which directs integration to the *ADE8* locus. We confirmed integration by polymerase chain reaction (PCR) and site-directed mutations by directly sequencing PCR products. Disruptions of *srb10* were created using pLS023, which is a two-step *URA3* disruption plasmid.

Recombinant protein expression

Recombinant His₆-Ste12 was produced by expression in baculovirus using the Bac-to-Bac system (Invitrogen) and purified using Ni²⁺ agarose according to the manufacturer's specifications (Qiagen). GST-Dig1 and GST-Dig2 were expressed and purified from *Escherichia coli* as described⁹. Srb10-Srb11 wild-type and D290A Srb10 complexes were purified from co-infected *Sf9* cell lysates as described¹⁵. We labelled yeast with [³²P]orthophosphate and carried out immunoprecipitations and tryptic phosphopeptide analysis as described⁶.

In vitro kinase assays

For assays with immunopurified haemagglutinin A (HA)-tagged Srb10, yeast were grown to mid-log phase in SC medium containing 2% glucose, collected and washed with sterile deionized water. Cells were resuspended in SLAD medium and grown for 4 h at 30 °C, collected, washed with sterile deionized water and stored at -70 °C. Immunoprecipitation of HA-Srb10 and subsequent kinase reactions were done as described¹⁵. We carried out kinase assays with recombinant Srb10-Srb11 complexes as described¹⁵, using roughly 100 ng of purified substrate protein for each assay.

Received 21 May; accepted 21 October 2002; doi:10.1038/nature01243.

1. Roberts, R. L. & Fink, G. R. Elements of a single MAP kinase cascade in *Saccharomyces cerevisiae* mediate two developmental programs in the same cell type: mating and invasive growth. *Genes Dev.* **8**, 2974–2985 (1994).
2. Cullen, P. J. & Sprague, G. F. Jr Glucose depletion causes haploid invasive growth in yeast. *Proc. Natl Acad. Sci. USA* **97**, 13619–13624 (2000).
3. Madhani, H. D. & Fink, G. R. Combinatorial control required for the specificity of yeast MAPK signaling. *Science* **275**, 1314–1317 (1997).
4. Morillon, A., Springer, M. & Lesage, P. Activation of the Kss1 invasive-filamentous growth pathway induces Ty1 transcription and retrotransposition in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **20**, 5766–5776 (2000).
5. Holstege, F. C. *et al.* Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* **95**, 717–728 (1998).
6. Hung, W., Olson, K. A., Breitkreutz, A. & Sadowski, I. Characterization of the basal and pheromone-stimulated phosphorylation states of Ste12p. *Eur. J. Biochem.* **245**, 241–251 (1997).
7. Boyle, W. J., van der Geer, P. & Hunter, T. Phosphopeptide mapping and phosphoamino acid analysis

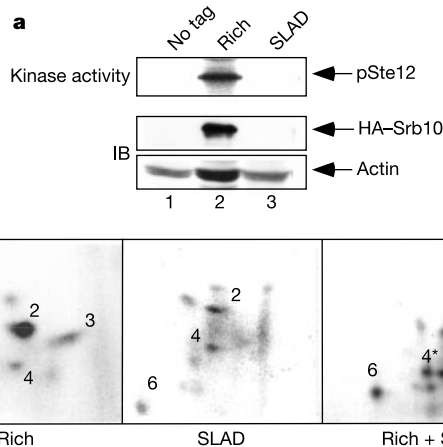


Figure 4 Nitrogen limitation inhibits Srb10 *in vivo*. **a**, Srb10 activity towards Ste12. Mid-log phase cultures of diploid yeast carrying integrated HA-tagged (lanes 2 and 3) or untagged *SRB10* (lane 1) were shifted from rich medium to nitrogen-limiting SLAD medium for 4 h. For kinase activity, immunopurified HA-Srb10 was used in *in vitro* kinase reactions containing recombinant Ste12 as substrate. For the immunoblot (IB), 50 μ g of protein extract was analysed with antibodies to actin, or was immunoprecipitated with antibodies to HA and immunoblotted with antibodies to Srb10. **b**, Tryptic phosphopeptides of wild-type Ste12 from diploid yeast grown in rich (left) and nitrogen-limiting (centre) conditions. Phosphopeptides from rich and nitrogen-limiting samples co-migrated (right). Phosphopeptide 4 resolves into two spots (4 and 4*) on co-migration, owing to a second Srb10-independent phosphorylation (not shown).

- by two-dimensional separation on thin-layer cellulose plates. *Methods Enzymol.* **201**, 110–149 (1991).
8. Pi, H., Chien, C. T. & Fields, S. Transcriptional activation upon pheromone stimulation mediated by a small domain of *Saccharomyces cerevisiae* Ste12p. *Mol. Cell. Biol.* **17**, 6410–6418 (1997).
 9. Olson, K. A. *et al.* Two regulators of Ste12p inhibit pheromone-responsive transcription by separate mechanisms. *Mol. Cell. Biol.* **20**, 4199–4209 (2000).
 10. Fields, S. & Herskowitz, I. Regulation by the yeast mating-type locus of STE12, a gene required for cell-type-specific expression. *Mol. Cell. Biol.* **7**, 3818–3821 (1987).
 11. Chang, Y. W., Howard, S. C., Budovskaya, Y. V., Rine, J. & Herman, P. K. The rye mutants identify a role for Ssn/Srb proteins of the RNA polymerase II holoenzyme during stationary phase entry in *Saccharomyces cerevisiae*. *Genetics* **157**, 17–26 (2001).
 12. Chi, Y. *et al.* Negative regulation of Gcn4 and Msn2 transcription factors by Srb10 cyclin-dependent kinase. *Genes Dev.* **15**, 1078–1092 (2001).
 13. Rohde, J. R., Trinh, J. & Sadowski, I. Multiple signals regulate *GAL* transcription in yeast. *Mol. Cell. Biol.* **20**, 3880–3886 (2000).
 14. Gimeno, C. J., Ljungdahl, P. O., Styles, C. A. & Fink, G. R. Unipolar cell divisions in the yeast *S. cerevisiae* lead to filamentous growth: regulation by starvation and RAS. *Cell* **68**, 1077–1090 (1992).

15. Hirst, M., Kabor, M. S., Kuriakose, N., Greenblatt, J. & Sadowski, I. *GAL4* is regulated by the RNA polymerase II holoenzyme-associated cyclin-dependent protein kinase SRB10/CDK8. *Mol. Cell* **3**, 673–678 (1999).

Acknowledgements We thank G. Sprague for plasmids; S. Hahn for antibodies against Srb10; and G. Mackie for comments. This research was supported by grants from the Canadian Institutes of Health Research (CIHR) and the National Cancer Institute of Canada (NCIC), with funds from the Canadian Cancer Society. C.N. was supported by a CIHR doctoral training award.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to I.S. (e-mail: sadowski@interchange.ubc.ca).

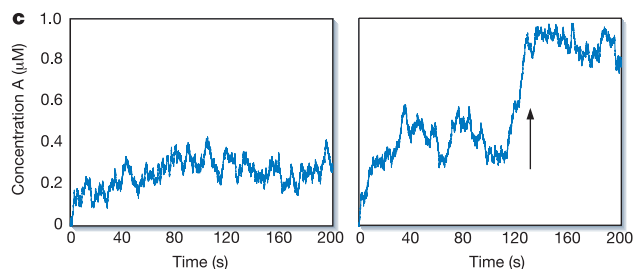
erratum

Control, exploitation and tolerance of intracellular noise

Christopher V. Rao, Denise M. Wolf & Adam P. Arkin

Nature **420**, 231–237 (2002).

In this Insight Review Article, the right panel of Fig. 3c incorrectly appeared blank. Figure 3c should have appeared as shown:



corrigendum

Altered performance of forest pests under atmospheres enriched by CO₂ and O₃

Kevin E. Percy, Caroline S. Awmack, Richard L. Lindroth, Mark E. Kubiske, Brian J. Kopper, J. G. Isebrands, Kurt S. Pregitzer, George R. Hendrey, Richard E. Dickson, Donald R. Zak, Elina Oksanen, Jaak Sober, Richard Harrington & David F. Karnosky

Nature **420**, 403–407 (2002).

In this Letter, the conversion to SI units led to several errors. On page 404, left column, lines 16 and 17, the values should read 10–15 and 30–40 nanolitres per litre. On page 405, right column, lines 3 and 4, the values should read 360 microlitres per litre and 36.0–38.8 nanolitres per litre. On page 407, left column, lines 3 and 4, the values should read 560 microlitres per litre, and 46.4 to 55.5 nanolitres per litre. The conclusions of the paper are not affected. □

New ideas for biotechnology

This week's trend—small sample size, and lots of them.

Hydra-Plus-One

Apogent Discoveries

www.apogentdiscoveries.com

Ready to drop

This is a 96- or 384-channel syringe/needle array combined with a single-channel non-contact dispenser capable of aspirating or dispensing volumes ranging from 100 nl to 220 μ l. The single-channel dispenser is suitable for distributing valuable recombinant protein using a single microtube or vial as the source. Applications include sitting drop, hanging drop and microbatch crystallization screens and low-volume PCR set-up.

IEF buffers

Amresco

www.amresco-inc.com

Positives and negatives for protein research

Amresco have added isoelectric focusing (IEF) buffers to their Pro-Pure line of products for proteomics research. They are described as convenient, ready-to-use concentrates that produce consistent results from run to run. The 20 $\times$ buffers are available in both anode and cathode versions.

Tissue culture flasks

Midwest Scientific

www.midsci.com

Unzip a cell culture

The new peel-off tissue culture flasks from Midwest are made from TPP and feature a 150 cm² growth surface. The removable cover allows direct access to the flask's growth surface for rapid cell retrieval. Also available is a version with 115 cm² growth surface with a barrier (15 mm high) for an almost square culture surface (10 $\times$ 11 cm). The cover material on the flask is a high-grade, non-toxic, multi-layer plastic film, with outstanding transparency, enabling microscopic examination the contents of the flask. The tear-off tab of the film is positioned near the neck and



Open tops: easy access to cells, by Midwest.



Hydra Plus-One: microdispensing for PCR and high-throughput crystallization screens.

marked 'Peel-off'. The flasks are made of high-grade, low-toxin polystyrene with side volume graduations and a white marking space. Only the growth surface of the flask is tissue culture treated, to avoid undesirable adhesion of cells on the side and cover surfaces.

Handbook of Chemistry

Advanced ChemTech Europe

www.advancedchemtech.com

Free reference book

The 2003–2004 edition of the *Handbook of Combinatorial, Organic and Peptide Chemistry*, is a 250-page volume offering practical tips and expanded technical notes on synthesis, cleavage and coupling techniques, as well as detailing Advanced Chemtech's line of more than 1,700 chemical products. These include synthesis resins and reagents, standard protected amino acids and hundreds of unusual amino acids. The handbook is available free from: www.peptide.com/request.htm.

Onto-Express 2.0

Open Channel Software

www.openchannelsoftware.com

Don't get sidetracked

Onto-Express is a bioinformatics tool designed to reduce the amount of time needed to interpret the biology of genetic profiles. Version 2.0 features a statistical algorithm designed to steer researchers away from paths that will not provide answers and towards those that will. The company claims that this reduces the time needed for genetic analysis from weeks down to minutes. A five-day free trial of the software is available via www.onto-express.com.

MyOne Streptavidin

Dynal Biotech

www.dynalbiotech.com

Well-behaved magnetic beads

This addition to the Dynabeads range is a 1- μ m, monosized magnetic bead that is coupled with streptavidin. The sedimentation rate of these beads is low, yet their magnetic mobility and binding capacity are high (3,000 pmol biotin or 20–25 μ g biotinylated IgG per mg beads). The binding capacity and smooth hydrophilic surface of the beads maximizes signal generation, resulting in a high signal-to-noise ratio. Iron content ensures equal behaviour of the beads within a magnetic field, improving target separation and optimizing assays.

Ettan DIGE

Amersham Biosciences

www.amershambiosciences.com

Multiple choice

Ettan DIGE is a platform for two-dimensional fluorescence difference gel electrophoresis (2-D DIGE). It enables researchers to use multiplexing—the separation of multiple samples on a single gel. The system consists of three components: CyDye DIGE fluors, the Typhoon variable-mode gel imager, and DeCyder differential analysis software. Up to three fluorescently labelled protein samples can be loaded onto the same gel. After electrophoresis, the gel is scanned and DeCyder software used to locate and analyse the protein spots. Statistical data are derived intra-gel rather than inter-gel, making the rapid routine detection of differences in expression down to less than 10% with over 95% confidence level.

SensiChip array system

Qiagen www.qiagen.com

High sensitivity without amplification

SensiChip arrays are designed to perform microarray analysis with as little as 1 µg total RNA without amplification steps, providing a true gene expression profile. Customized SensiChip DNA arrays are made to the user's specifications and spotted with 70-mer oligos, which combine the specificity of oligo probes with the sensitivity of cDNA probes. Qiagen also offers the SensiChip human kinase DNA array bar, which contains six separate arrays carrying a comprehensive set of human kinase and phosphatase genes.

sciPROOF

Sciproof LLC www.sciproof.com

Prepare and edit scientific manuscripts

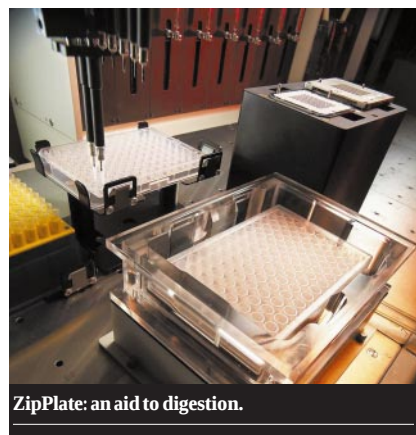
This software program can be integrated with Microsoft Word and used to streamline the scientific writing process. sciPROOF verifies spelling, style, acronyms and Greek symbols. There is also a glossary and users can check definitions and references via a direct link to the US National Library of Medicine, without leaving the Word document. Formats and styles can also be integrated and proofread within the document.

ZipPlate

Millipore www.millipore.com

Sample processing in a jiffy

The ZipPlate Micro-SPE Plate is a 96-well sample preparation tool that performs both in-gel digestion and sample preparation prior to mass spectrometry. Most laboratories currently digest protein-containing gel slices in a centrifuge tube and then purify the resulting peptides using a pipette tip or other purification device. The ZipPlate system has 300 nl of C18 media immobilized at the bottom of each well, so samples can be digested, desalted and concentrated without being transferred between devices. As a result, up to 96 samples can be processed in 7 hours, compared to 16 hours using current methods.



ZipPlate: an aid to digestion.



SensiChip: microarray analysis with as little as 1 µg RNA.

384-well microplates

Porvair Sciences www.porvairsciences.com

Image conscious

The standard range of 384-well microplates from Porvair Sciences is designed to enhance assay measurements and eliminate automation stoppages caused by variations in plate dimensions. A repeatable plate flatness of ±0.10 mm increases precision and eliminates misreads caused by the consistent focal plane of imagers. The plates are available in black, white or clear polystyrene and have been optimized for fluorescence, luminescence/scintillation and ELISA/turbidity measurements respectively.

Sealing tapes and membranes

Nunc www.nuncbrand.com

For use with MicroWell plates

Nunc's range of sealing tapes provide an effective seal on polypropylene, polystyrene and polycarbonate MicroWell plates, minimizing evaporation and well-to-well cross-contamination. Compatible with aqueous solutions and organic solvents, the tapes are also tested for use between -80° and +100° C. A DNase and RNase free version is available for thermal cycling applications. Optical clarity allows reading through the tape. All tapes are suitable for long-term storage. Applications include general assays, genomics, compound and library storage, bio-analytical assays, high-throughput screening, PCR and drug discovery.

Coded tube readers

Micronic www.micronic.com

Speedy sample scanning

These readers use a single-life vision system camera to take a visual image of the tube code that is then translated into a unique number. HTM readers can scan a rack of 96 codes in less than a second, facilitating high-speed movement and audit-traceable tracking of laboratory samples. For less demanding applications, the TM reader uses a standard flatbed scanner to scan and interpret 96 codes in under 20 seconds. Finally, for tracking

single tubes, the STM reader, which is available in either stand-alone or in-line configurations, can scan and interpret the codes on single tubes in less than 0.5 seconds. Readers are available for both TraXis and Tracker-DM coded tubes.

E. coli O157 array

MWG www.THE-MWG.com

Complete genome array

This array from MWG features 6,336 oligonucleotides representing 14,982 genes covering the complete genomes of three different *Escherichia coli* strains: 4,288 from the non-pathogenic strain K12, 5,336 from the pathogenic O157:H7 (EDL 933) and 5,358 from O157:H7 (RIMD0509952). The last two strains are virulent pathogens that cause severe damage to the lining of the intestine. *E. coli* O157:H7 has been found to possess more than 1,000 genes lacking in non-pathogenic K12, which may contribute to the virulence of the former. This array allows the creation of genomic tools to differentiate *E. coli* infections more easily and to identify potentially contaminated food, as well as aiding research into novel antibiotic drugs and vaccines.

PowerPlex ES

Promega www.promega.com

System solution for forensics

PowerPlex ES is a multiplex STR system for use in DNA typing, including forensic DNA analysis, human identity testing, tissue culture strain identification and paternity testing. It amplifies the seven ENFSI-approved STR loci, amelogenin and the highly polymorphic SE33 (ATBP2) locus in a single amplification. The system provides a matching probability of one in 5.914×10^{10} and a power of exclusion of 0.999946. Allele determination is simplified by the inclusion of an allelic ladder for all nine loci, including most of the frequently occurring microvariants.

These notes are compiled in the Nature office from information provided by the manufacturers. Press releases to be considered for publication can be sent to newproducts@nature.com.

Contacts

Publisher: Ben Crowe

Editor: Paul Smaglik

Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Senior European Sales Manager:
Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)

Andy Douglas (4975)

Frank Phelan (4944)

Netherlands/ Italy/ Iberia:

Evelina Rubio Hakansson (4973)

Scandinavia: Silje Opstrup (4994)

France/ Belgium:

Amelie Pequignot (4974)

Production Manager:

 Billie Franklin

To send materials use London
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

International

Advertising Coordinator:

Hind Berrada (4935)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Office

Germany/ Austria/ Switzerland:

Patrick Phelan, Odo Wulffen

Tel + 49 89 54 90 57 11/-2

Fax + 49 89 54 90 57 20

e-mail: p.phelan@nature.com

o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103

e-mail: naturejobs@natureny.com

US Sales Manager:

 Peter Bless

US Advertising Coordinator:

Linda Adam

Japan Head Office, Tokyo

MG Ichigaya Building (5F),
19-1 Harakatomachi,

Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

Asia-Pacific Sales Director:

Hideki Watanabe

e-mail: h.watanabe@naturejp.com

naturejobs

Sweet dreams...

Last month *Man of La Mancha* landed on Broadway, just a few days after a report on creating a European Research Area hit the Internet (www.euroscience.org). These seemingly unrelated events in fact have the same central element — *The Impossible Dream*. The song is the dramatic fulcrum on which the musical, based on Cervantes' novel *Don Quixote*, revolves. But the song's title could also be seen as the premise or promise of the European Research Area (ERA).

The ERA aims to provide easy access to well-funded scientific jobs across Europe for every European scientist, regardless of national affiliation. In some respects, the mobility of scientific employment in Europe hinges on the ERA's success in much the same way that the musical depends on its showcase song.

But right now, the ERA remains a dream because it is just "an empty shell", according to the Euroscience report, which was based on a November conference sponsored by the European Commission. The two biggest obstacles preventing the ERA from becoming a reality are low funding for science and technology, and widely varying employment policies among the European nations, the report says.

The report acknowledges that it can't do anything about the funding. But it does present a model for doing something about mobility. It recommends setting up a European Research Contract Managing Office, which would look after issues such as personnel grants awarded within Europe for integrated projects or network of excellence.

Only time will tell if such an office warrants Quixote's noble optimism expressed in *The Impossible Dream*, or whether it will ultimately deserve our cynical scorn for tilting at unyielding bureaucratic windmills.

Paul Smaglik

Naturejobs editor



Contents

POSTDOCS & STUDENTS

Banding together for a
better future

p194

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

SCIENTIFIC ANNOUNCEMENTS

SCIENTIFIC EVENTS

Seeking strength in numbers

Postdocs are vital to scientific research, but often miss out on the benefits available to permanent staff. Now they are banding together to improve their situation, says Karen Kreeger.

A decade ago, riled by low stipends and poor healthcare benefits, postdocs at Johns Hopkins Medical School in Baltimore, Maryland, banded together to do something about them. Since then, dozens of similar grass-roots postdoc associations have emerged at other universities, along with organizations for postdocs that are funded and staffed by the institutions themselves. How much have these groups achieved and how effective are they in bettering the postdoc's lot?

The Johns Hopkins postdoc association tries to answer these questions through regular surveys. The most recent, in late 2001, canvassed the 1,235 postdocs working in the medical school. It found that 60% would encourage or strongly encourage others to take up a fellowship at Johns Hopkins. This is pretty good, considers Carol Manahan, president of the postdoctoral association for the Johns Hopkins School of Medicine. And she confesses herself surprised at this level of satisfaction.

The respondents' top concern was salaries. At the time of the survey, Johns Hopkins had yet to achieve the salary standards for postdocs set out in guidelines from the National Institutes of Health, but it has since implemented the recommended minimum salaries (see *Naturejobs* 5; 10 January 2002). The next four most pressing concerns were, in order, future job placement, parking, childcare and healthcare benefits. Manahan says that the association is now working to obtain more comprehensive healthcare cover.

committees set up by the council, "the biggest one we have going on now is the foreign postdoc committee". Because foreign postdocs tend to be "isolated in a lab", says Pappas, one of the council's goals is to bring them together socially, whether that is by hosting pot-luck dinners with relevant administrative departments or by organizing a soccer team.

WIDENING HORIZONS

Postdocs at the University of California (UC) have taken their cause state-wide. In late 2001, they set up the UC Council of Postdoctoral Scholars to tackle problems that could not be solved by any single campus. Take job descriptions. On the ten UC campuses "there were as many as eight different job titles for postdocs, with the accompanying differences in benefits and salaries", says Raymond Clark, chair of the council and a postdoc at UC, San Diego. After the council's intervention there are now just two. The next issue in the council's sights is a uniform benefits and salary structure for postdocs throughout the university.

Last spring, postdocs even went national, forming the National Postdoctoral Association to advocate best-practice policies. "It was very much a grassroots effort," says Orfeu Buxton, a postdoc at the University of Chicago who is on the association's steering committee together with Manahan. "There are certain problems that are national in scope, and at a local level intractable," explains Buxton. He cites as examples the lack of clarity in the

US tax code regarding postdocs and the lack of consistency between funding agencies, which can generate disparities in salaries between, and even within, labs.

Since March 2002, the National Postdoctoral Association has established collaborations with the Association of American Universities and Sigma Xi among others. "We're not proposing anything socialist," says Buxton wryly. "It's just that some rationality would help. So rather than wailing into the wind, there's a huge amount of action going on about postdocs, and we could add a national focusing voice."

Karen Kreeger is a freelance science writer based in Philadelphia.



John Pappas
Foreign postdocs tend to be isolated in a lab and our council helps them get together socially.

POSTDOCS FROM OVERSEAS

Half the postdocs at Johns Hopkins come from outside the United States, and the survey revealed that they emphasize different issues from their US counterparts. As well as the obvious problems of a language barrier and visa difficulties, postdocs from overseas seemed to be less well paid on average, and they put communication difficulties and conflicts with their principal investigator higher on their lists of problems.

The concerns of foreign postdocs are also high on the agenda of the Biomedical Postdoctoral Council at the University of Pennsylvania School of Medicine. John Pappas, a postdoc in the microbiology department and co-chairman of the council, says that, of the issue-led

Web links

National Postdoctoral Association

♦ www.nationalpostdoc.org

Johns Hopkins School of Medicine Postdoctoral Association

♦ www.med.jhu.edu/jhpda

National Postdoctoral Association Steering Committee

♦ www.nationalpostdoc.org

University of California, Berkeley, Postdoctoral Association

♦ postdoc.berkeley.edu/content_home.html

University of California Council of Postdoctoral Scholars

♦ www.psa.ucsd.edu/policies

University of Chicago, biological sciences postdoc association board

♦ pda.bsd.uchicago.edu

University of Pennsylvania biomedical postdoctoral programmes

♦ www.med.upenn.edu/postdoc